The Folly of Reason

Mathematical logic and its paradoxes

Everything is mathematical

The Folly of Reason

The Folly of Reason

The Folly of Reason

Mathematical logic and its paradoxes

Javier Fresán

Everything is mathematical

Published by RBA Coleccionables, S.A.
c/o Hothouse Developments Ltd
91 Brick Lane, London, E1 6QL

Localisation: Windmill Books Ltd.
Photography: gettyimages, agefotostock, Corbis

ISSN: 2050-649X

Printed in Spain

For José Antonio Pascual
and Rosa Navarro Durán

Contents

Preface

A couple start to argue: "You always disagree with me. You always say the opposite," says the wife. "That's not true," replies the husband. "You see? You've just proved it yourself," she attacks again. "You're right, dear. I always say the opposite," admits the man, in an attempt to put an end to the argument. "Well, it must be really bad if you even admit it yourself!" shouts the woman, walking out the door. Such scenes are everyday occurrences, even in the best families. If the mathematician and philosopher Bertrand Russell had never been through such encounters, it is certain he wouldn't have been married four times. His arguments, however, would have ended in a different way. After the "You've just proved it yourself," Russell would have been silent for a few moments and, perhaps after making a comment such as, "What you say is very interesting," would have locked himself in his study.

To do what? To think about statements that speak of themselves, about the true and the false, until he came up with a paradox that would throw doubt on the supposition that the mathematics of the previous two thousand years was the most perfect accomplishment of intellect. Russell's paradox is one of this book's protagonists, but, as it didn't seem right to start the story in the middle, I first had to tell of how the discovery of non-Euclidean geometries radically changed the axiomatic method, and how the contradictions that put an end to the English philosopher's "glad confident mornings" had their roots in a tradition that goes back at least to Epimenides of Crete. Russell's paradox, however, would have been nothing more than an oddity had it not been for the answers that it threw up. We'll first look at the solution provided by David Hilbert, one of the most brilliant men of his time, who for 30 years kept alive the hope that mathematics would once again be utterly reliable. That's what young Kurt Gödel would have liked to prove, but instead he discovered that even in mathematics there are some truths that are not provable.

Since he discreetly announced them at a conference in September 1930, Gödel's incompleteness theorems have fascinated both mathematicians and humanists. Some have chosen to believe they were the defeat of reason in the battle where they were intended to dominate. Others saw them as irrefutable proof of the superiority of human beings over machines. But only those who truly understood the formalism in Gödel's articles were able to channel their logic towards new territories. Precisely by reinterpreting the incompleteness theorems, the man who had deciphered the Nazis' diabolical cryptography, the brilliant Alan Turing, was able to envisage the forerunners of our computers. All of this, and many things more, are the subject of

this book, which does not limit itself to the 'ones or zeros' of Turing machines, but tries to go one step furtherto describe the nuanced world of one of the most recent incarnations of the dream of reason: fuzzy logic.

I would like to thank the publishers for their invitation to write this book. It was, in fact, the words 'narratives of popularisation' hidden away in one of the emails I exchanged with the editor that gave me the idea of beginning each chapter with some light touches of literature. Without the tales of that 21st-century Scheherazade, my friend Laura Casielles, I would never have been able to relate fuzzy logic to the desserts served in a Japanese restaurant. The beginning of Chapter 5 owes a lot to the fascination for Alan Turing felt by Patricia Fernández de Lis. The work has been improved thanks to the meticulous reports with which Jesús Fresán, David Garcés, Miguel Hernaiz, Victoria Ley Vega de Seoane, Javier Martínez and Luz Rello immediately responded to my dispatch of the first drafts. I am very grateful to all of them, as I am to María Aguirre Roquero, Luis Azcárate, Noel Garrido, Geno Galarza, María Ángeles Leal, Carlos Madrid, José María Mateos, Guillermo Rey, Roberto Rubio, María José Soler, Lucas Sánchez and Mikel Tamayo for their valuable suggestions.

Chapter 1

The Axiomatic Method

Since the Greeks, whoever says mathematics needs proof.
Nicholas Bourbaki

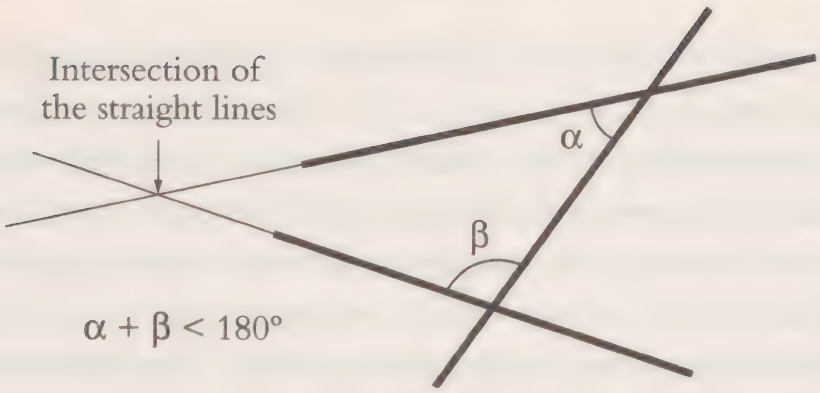
The enthusiasm with which lawyer Taurinus had ripped open the envelope, without even looking round for a letter-opener, began to turn to disappointment as he read through the two pages of dense handwriting he had received that November morning in 1824. The letter contained the reply from Carl Friedrich Gauss to the announcement of a discovery of extraordinary importance – a proof of Euclid's fifth postulate.

Gauss was now almost 50 years old, and there was not one branch of physics or maths in which he had not made a huge impact through countless contributions which earned him the title of *princeps mathematicorum*, the 'Prince of Mathematics'. However, none of his works had taken on the burning question of the time. Was the fifth postulate true? Could one and only one parallel line be drawn through a point outside a given straight line? Answering this question meant – to a certain extent – answering the question 'What form has the world?'

The story of Euclid and the work that sets out his ideas – *Elements of Geometry* – goes back to 300 BC. At this time the Greek mathematician, about whom we know little more than his name, had written a manual on geometry setting out the corpus of knowledge that had been passed down orally through the centuries by the Pythagoreans and the followers of Plato. But, contrary to what was proclaimed over the doorway to that philosopher's Academia, "Let no one ignorant of geometry enter here," Euclid's *Elements* were used by readers to learn the mathematics of shapes starting from the most basic principles. With the dual purpose of smoothing the path for his students and injecting some order and rigour into the scientific tradition, Euclid began his treatise with a series of definitions and axioms that allowed patient readers to deduce any of the hundreds of further propositions included in the book. Perhaps no other pedagogical work has had such radical consequences in more than two thousand years.

Euclid painted by Raphael in The School of Athens. *The Greek mathematician, shown taking measurements with a pair of compasses, is surrounded by his followers.*

Dictionaries define axioms as self-evident statements that are admitted without the need for justification. In this sense, they do not form part of a legacy of a culture but are conclusions that a human living outside of civilisation would reach independently. Euclid differentiated between common notions and postulates. Axioms of the type: "If two things are equal to a third thing, then they are equal to each other," work just the same when talking about regular polygons as when talking about the gods. However, postulates are specific to geometry. For the wise man from Alexandria, five of these pillars were enough for the *Elements*. The first three state that a straight line can be drawn passing through two points, any straight line segment can be extended, and any circle of any radius can be drawn around an arbitrary centre; the fourth states that all right angles are equal. According to the fifth, the one that had kept Taurinus busy for months, if a straight line intersects another two lines in such a way that the interior angles on the same side add up to less than 180°, then the two other lines will intersect at a point situated in the same half of the plane on which the angles are situated.



The point where the two straight lines intersect is in the same half of the plane as the angles.

The modern reader's first impression is probably no different from that of Euclid's contemporaries, that the fifth postulate isn't as self-evident as the previous ones and that you need a pen and paper to understand it. That's why the geometricians soon began to argue about its status as an axiom and tried to prove it by using the others. Even though all their attempts turned out to be in

A CONVERSATION FROM THE FILM *AGORA* (ALEJANDRO AMENÁBAR/MATEO GIL, 2009)

Hypatia: Synesius, what is Euclid's first rule?

Synesius: Why the question?

Hypatia: Just answer me.

Synesius: "If two things are equal to a third thing, then they are equal to each other."

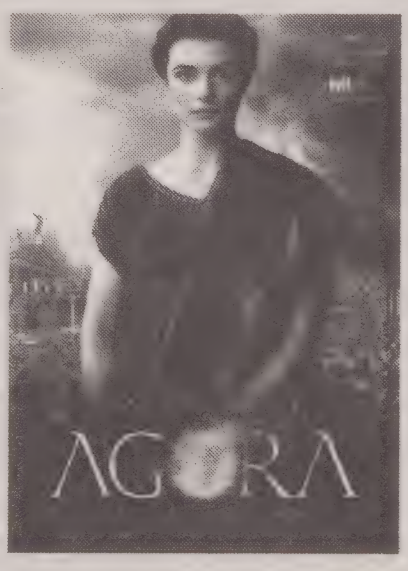
Hypatia: Good. Now, are you both not similar to me?

Synesius: Yes.

Hypatia: And you, Orestes?

Orestes: Yes

Hypatia: Now I am actually saying this to everybody here in this room: more things unite us than divide us, and whatever may be going on in the streets we are brothers. We are brothers. I want you to remember that brawls are for slaves and riffraff.



A poster advertising Agora, *a film whose protagonist is Hypatia of Alexandria.*

vain, on the way, other expressions equivalent to the fifth began to appear that helped to understand its consequences. The most famous ones stated that the angles of a triangle add up to 180° and that only one parallel line can be drawn through a point lying outside a straight line. In this and other similar forms, the doubts over whether what came to be called the *parallel postulate* was really independent of the others, or whether in fact some ingenious argument would make it possible to eliminate it from the list of axioms. Such doubt would outlive all the Classical Greek, Arab and Renaissance commentators on the *Elements*.

So how surprised Franz Adolph Taurinus must have been on that November morning on learning that, instead of covering him with glory for managing to go further than the best minds in history, the great Gauss confessed that, after 30 years wondering about whether Euclid had told the whole the truth, he had become sure that a geometry that did not comply with the fifth postulate was possible. But this new non-Euclidean form had to be kept secret until all the details of a theorem were ready because it would contradict an image of the nature perpetuated for two millennia. It would not have been welcomed by those who supported the idea that the triangles and circles with which the book of nature was written were just as Euclid had described them. Just like Aristotle for the Scholastic philosophers, Euclid was not just a man but an almost sacred figure of knowledge.

From non-Euclidean geometry to relativity

The above could make up the beginning of a story based on actual events. In its next chapter, Gauss (1777-1855) would measure the triangle formed by the tops of three mountains in Germany so as to decide once and for all whether the geometry of space was or was not Euclidean. During the story the Prince of Mathematics would be joined by other characters such as the Hungarian János Bolyai (1802-1860) and the Russian Nikolai Lobachevski (1792-1856), who did not take so many precautions when proclaiming their discoveries.

In aristocratic salons, scientists from all over Europe would charm a captive audience by showing them models of strange surfaces, in which the angles of the triangles always came to less than 180°. And someone would interrupt the magic show to shout "Euclid is dead!" while the more conservative among them would react with horror: "No man can serve two masters: if Euclidean geometry is the true geometry,

then non-Euclidean geometry must be deleted from the list of sciences and placed next to alchemy and astrology."[1]

That is not the story line of this book. The tale we are going to tell you also begins with the discovery of the new geometries and has an even more unexpected outcome, out of which computers emerge and the first experiments on artificial intelligence get under way. Apart from the windows that were opening to other worlds, the most important implication of the existence of non-Euclidean models was of a philosophical nature. Euclid had chosen those axioms and not others because their truth was self evident. However, from the moment it became apparent that an infinite number of parallel lines could be drawn through a point on some surfaces, while other bodies of a different form had none at all, the question of which of the axioms was true lost its meaning. Why should the parallel postulate be any more true than its negations? In reality, how valid the different postulates were simply depended on the objects chosen for investigation.

One of those who best knew how to exploit the field once the battle was over was Albert Einstein (1879-1955) who, thanks to one of the non-Euclidean geometries, was able to solve a problem that had kept even Isaac Newton (1643-1727) himself awake at night. According to the law of universal gravitation, which was established by the English scientist in 1685, two bodies attract each other with a force that is greater in proportion to the product of their masses and lesser in proportion to the square of the distance that separates them. This principle enables the movement of the planets and fruits falling from a tree to be explained in a unified manner, but did not answer a fundamental question: how can the Earth exert force on the Moon if they are separated by almost 400,000 kilometres? Action at a distance was then an invention attributed to alchemists, and so could in no way be accepted by mainstream thinking as being the guarantor of equilibrium in the Universe.

To avoid the issue, they looked to Greek mythology and resuscitated ether, an ephemeral substance that filled the gaps in a vacuum and through which the attraction of gravity would be transmitted from one body to another. However, a number of experiments threw doubt on whether anything of the sort actually existed.

---

[1] A letter from Gottlob Frege (1848-1925) to David Hilbert (1862-1943).

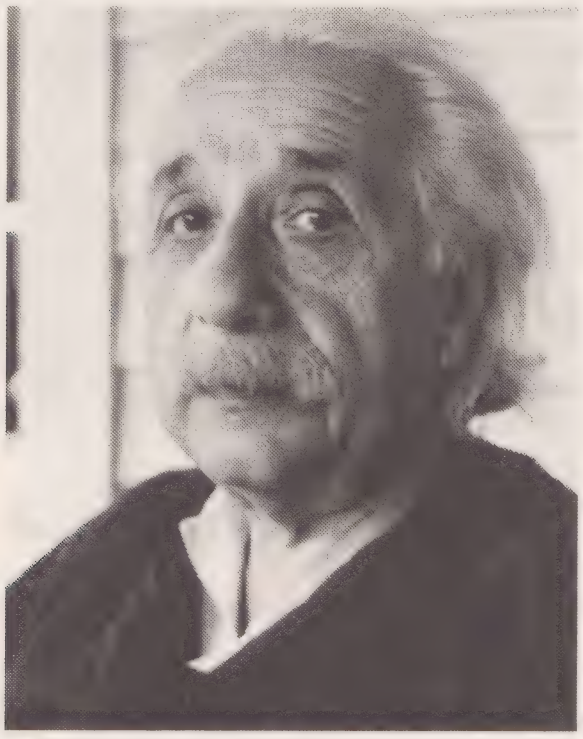
And that's where Einstein came in. Anyone can imagine what happens to a sheet held by two people when a heavy ball is dropped into the centre, but it took the imagination of this brilliant patent office clerk in Berne to come up with the idea that the same thing happened to planets – any body in fact – in space. A body with such a great mass as the Earth deforms the vacuum around it, and gravity is nothing more than that curvature in the Universe. In the same way that a marble thrown into the sheet deformed by the ball will immediately be attracted towards the centre, when a body is left free near to the surface of the earth, the slope that the Earth has created in space will make it fall. If the body is further away and on the move, such as the Moon, for instance, the deformation of the Universe will not make it rush towards our planet, but rather, the body will stay in orbit around it. Well then, it turns out that in the geometry we use to identify gravity with a curvature in space, Euclid's fifth postulate is not true.

A graphical representation of the deformation of space caused by the gravitational force of the Earth.

Einstein, however, was not in the least bothered that his theory of relativity had banished the dream of a Euclidean cosmos, because over time he had come to the realisation that geometry was a formal concept. In the first chapter of *Relativity: The Special and General Theory*, a popularisation of his research published in 1920, Einstein explains that geometry is based on a handful of concepts such

as 'point', 'plane' and 'straight line', of which we have a defined image, and on a series of simple propositions, the axioms, which seem true to us when we interpret them according to the idea we've used to form the objects they refer to. Starting from these basic principles, all the other propositions are proven by following a logical process of deduction, so that if we admit that the reasonings used are correct, the truth of the result rests only on the truth of the premises. So far so good. To answer the question of what form the world has we need to know if the five postulates are true or not. But not only is that question impossible to solve by geometric methods, but it also lacks sense.

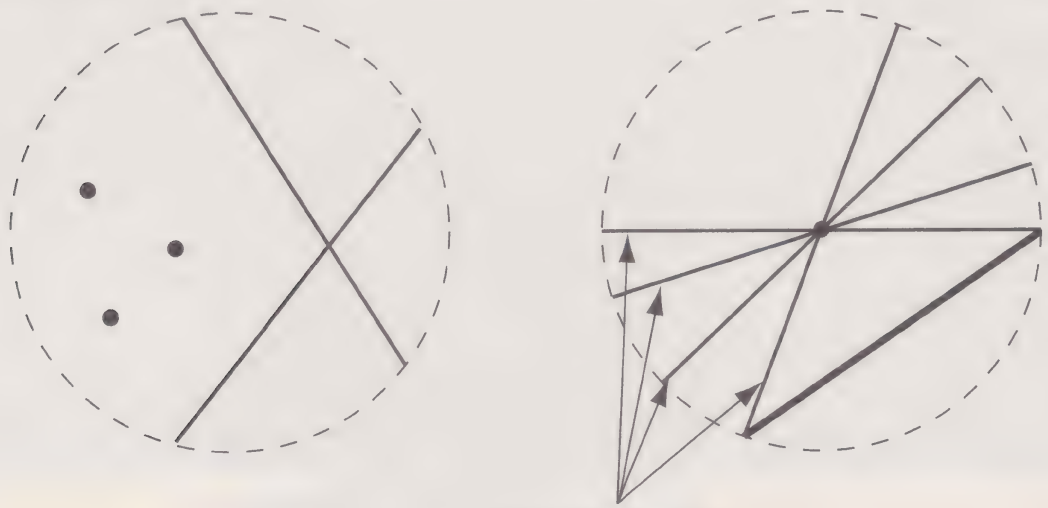
It's a waste of time trying to determine – Einstein continues – if it is true that only one straight line passes through two points. All we know is that geometry deals with things that are called 'point' and 'straight line', whose relationship is the following: Two different 'points' determine one and only one 'straight line'. For a discussion on the truth of axioms to make sense, we first have to establish a correspondence with reality. If every time Euclid says 'point' and 'straight line', we think of what everyone understands by those words, then the axiom "a straight line can be drawn passing through two points" becomes a tangible statement. And then we can prove experimentally, so to speak, whether it is true



Taken around 1950, one of the last photographs of Albert Einstein.

or false. However, there is nothing to indicate that geometry is a translation of the language of everyday objects, but rather just a set of abstract relationships between made-up concepts.

Let's look at an example that first appeared in two papers by the Italian geometer Eugenio Beltrami (1835–1900). The space in which objects are situated will be, from now on, the interior of a circle, without including the edge, and the correspondence we are going to propose is quite simple. When Euclid says 'point', we shall think of the points inside the circle, and when he says 'straight line', we shall imagine the segments that start and end on the edge of the circle. With this translation, two 'points' determine just one 'straight line' and, therefore, Euclid's first postulate is true. To see what happens with the fifth postulate, remember that two 'straight lines' are parallel if they never intersect each other. Let's take any 'point' within the circle, for example, the centre, and an arbitrary 'straight line'. On joining the 'point' to the ends of the line segment, we get two 'straight lines', which pass through it and which are parallel to the first one, as the hypothetical points of intersection are on the edge, which does not belong to the space! Therefore, in Beltrami's model, the parallel postulate is not true.

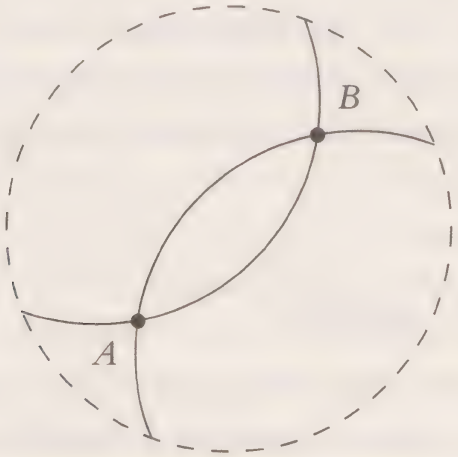


Eugenio Beltrami's non-Euclidean model.

Note that in the previous paragraphs the words 'point' and 'straight line' are sometimes in inverted commas and sometimes they aren't. In this way we wanted to distinguish the abstract concepts of 'point' and 'straight line', which could have

very different interpretations, from the real points and straight lines which gave rise to them. If anyone thinks we have cheated by presenting this non-Euclidean toy, maybe a bit of biology would make them change their minds because in the best of cases, human sight only reaches a few kilometres. As a consequence of this limit, all straight lines that intersect beyond the edge of the circle are the same to our eyes, and what we see around us fits in reasonably well with the image of the Italian geometrist. After all, what difference can a European see between two straight lines that intersect in New York and another two that intersects in Los Angeles? The small world of the human being is not Euclidean. But there are more things on heaven and earth than are dreamt of in his philosophy...

What we need to point out here is that Beltrami's proposal is only one arbitrary choice between the many that are possible. In the space itself we could have called the circle's arcs 'straight lines' and then the first postulate would not be verified, as there is no limit to the ways of joining two points in this way. To determine a circumference completely three points are necessary, and it is precisely that freedom to choose the third that prevents the axiom from being fulfilled. If some models satisfy the first postulate and some don't, the fact that just one 'straight line' passes through two 'points' must necessarily depend on the meaning given to the concepts 'point' and 'straight line'. To puzzle over the truth of it is as pointless as to try to discover if there is any truth in the prophesy "In the year *A* shall be born *B*", where the reader can replace, if they want, *A* and *B* by sufficiently vague expressions.



A space in which two different lines join two points A *and* B *and in which Euclid's first postulate is not verified.*

That is what we meant before when we said that Einstein was aware that geometry is a formal construction. Nevertheless, he was not interested in logical relationships between concepts, but in the specific question of explaining action at a distance without having to resort to ether. His 'points' were points in space, localised by means of coordinates which said where they are and at what instant we look at what is happening there. His 'straight lines' were the fastest paths between two points, which are the paths that a ray of light follows. If the means that Einstein needed to make his predictions about the nature of space was a negation of the parallel postulate, why shouldn't he use it? In May 1919, four years after Einstein identified gravity with a curvature of the Universe, an expedition to the island of Principe, off the coast of west Africa, managed to observe the light from stars situated near the Sun bending around it during an eclipse. It was this type of experimental verification, together with the consonant theoretical results, that could tell us something about the validity of relativity, and not the fact that a non-Euclidean geometry had been necessary to formulate the laws.

Naturally, it did not occur to Euclid that his 'points' and his 'straight lines' could be replaced by almost anything when he began to write the *Elements*. For Euclid it would have been no more than a language game, a provocation, because for him the components of geometry over and above everything else had a physical meaning. Proof of that is the way in which he enunciates his axioms, which say that, given two points, a straight line can be drawn joining them, and not that for any pair of 'points' there exists just one 'straight line' containing them, as we have the tendency to interpret it. With the subtle jump from points to 'points' and from the 'can be drawn' to the 'there exists', which separates one version from the other, geometry became abstract and mathematical logic was born.

The new axiomatic systems

The first change that this revolution required was to reconsider the concept of axiom, as there was no sense in looking for 'evident truths'. After the birth of non-Euclidean geometries, an axiom would be no more than a statement that was placed by convention in the basis of a theory so as to enable theorems to be deduced from it. The marvellous thing about a language is that it allows words to be combined as we like, and as long as a few rules are followed, the

BASIC LOGIC SYMBOLS

One way to remember the structure of *modus ponens* and of *modus tollens* is to show them in formulae in which a line separates the premises from the conclusion. If we use $\neg A$ and $\neg B$ to denote the negations of A and of B, in other words, the propositions that state the opposite, then the *modus ponens* and the *modus tollens* correspond to the following diagrams:

$$\begin{array}{l} A \to B \\ A \\ \hline B \end{array} \qquad \text{and} \qquad \begin{array}{l} A \to B \\ \neg B \\ \hline \neg A \end{array}$$

person we're talking to will be able to understand a sentence even though it might be the first time it has ever been articulated. However, when we invent a word we are obliged to explain its meaning to the listener, and the word is not likely to survive if there is no agreement on its usefulness or its beauty when referring to things. With logic, something similar happens: a proposition cannot be proven starting from square one, but rather we first need to provide it with some principles that everyone is aware of, along with some rules of deduction or inference which we will use to get beyond the axioms.

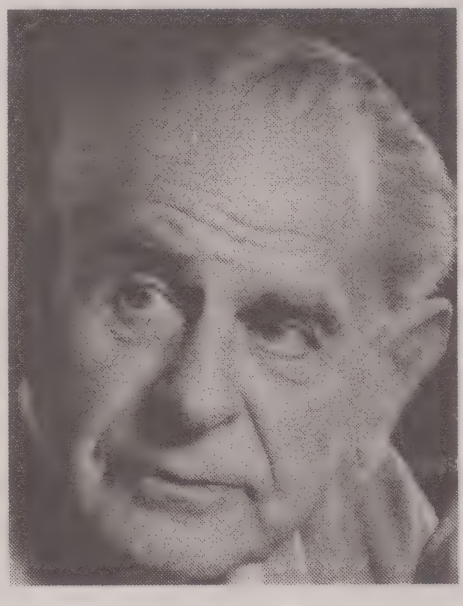
A classic example of these rules is *modus ponendo ponens*, or simply *modus ponens*, which in Latin means 'the way that affirms by affirming' and which consists of deducing from the implication 'If A, then B' and from the verification of A, that B is true. We must again emphasise that the meaning of the rules of inference is, like that of axioms, purely formal. Hence the deduction that 'All men can fly. Icarus is a man, so he can fly' is correct, while 'If it rains, the pavement gets wet. The pavement is wet, so it has rained' cannot be considered a valid inference. Although the notion of the pavement being wet because of the rain is reasonable, and that of the man flying is completely absurd, the first deduction is correct whereas cause and effect have been mixed up in the second one. Rain implies that the pavement is wet, but the pavement being wet does not imply necessarily that it has rained. For instance, a shopkeeper might have thrown buckets of water over it. There is also *modus tollendo tollens*, or just *modus tollens,* which is a 'way of denying by denying', this time consisting of deducing from the implication 'If A then B' and from the fact that B is not verified, then

neither is A, as in the reasoning: 'Of what one doesn't know one must not speak. If I speak it is because I know'.

In general, a rule of deduction is only valid when its conclusion is true irrespective of how its premises are interpreted. So, the inference 'If P and Q, then R' is correct whatever the meaning we give to P, Q and R if, as long as P

MODUS TOLLENS AND 'FALSIFICATION'

According to the philosopher Karl Popper (1902–1994), *modus tollens* is the only legitimate deduction in natural science. To explain any phenomenon, the scientific method, which Popper called 'hypothetical-deductive', states hypotheses and design experiments that enable them to be verified. If from a hypothesis *H* an observable consequence *O* follows, which we verify repeatedly in the laboratory, then H becomes a scientific law. But, unless we can verify one by one all the situations to which our hypothesis is applied, we will never be absolutely sure that it is fulfilled. For us to be sure that all swans are white they would all have to examined one by one all around the world for ever, but it would be enough to come across just one black one, as happened to the first explorers of Australia, to refute the hypothesis. This principle is known as **'falsification'.** It is nothing more than a *modus tollens*: 'If the hypothesis *H* is true, then consequence *O* will follow. As the opposite of *O* is observed, we deduce that *H* is false'.



The philosopher Karl Popper in the 1980s.

and Q are verified together, R is also true. We again find ourselves with a formal criterion that implies, for example, that the deduction 'If zero is different from one, and one is equal to zero, then you are my father' is valid. As in none of the possible worlds zero is equal to one and different from one at the same time, the hypotheses are never verified, and there is nothing to prove. The Scholastics

had realised this and very aptly coined the expression *ex contradictione sequitur quodlibet*, in other words, 'From a contradiction anything follows'.

Now that we know what axioms and the rules of inference are, we are ready to specify the terms 'theory', 'demonstration' and 'theorem', which have turned up in the previous pages with meanings that have been, more or less, intuitive. A demonstration, which sometimes we'll call proof, is the process that enables new results to be obtained by applying the rules of inference to axioms. In practice, it consists of a finite series of affirmations, also called statements, of which the first must necessarily be an axiom (in mathematics there are no clean slates!), and each of the following ones can be an axiom or can be deduced from the preceding affirmations by applying the rules of inference. The last statement of a demonstration is called a theorem, and a theory is a collection of axioms, of rules of inference and of all the theorems that can be proven by them based on the axioms. On some occasions, instead of 'theory' we shall say 'axiomatic system'.

Up to now we have focused on Euclidean geometry, which is the theory made up of the five postulates of the *Elements*, of the rules of deduction such as "If two things are equal to a third thing, then they are equal to each other," and of all the theorems on circles, triangles and polyhedra that the reader can possibly imagine. We have also dealt with non-Euclidean geometries, which share with it the first four axioms and add a negation of the fifth. But the authentic protagonist of this book is arithmetic, a theory that deals with numbers used for counting, which we mathematicians *officially* call 'natural'.

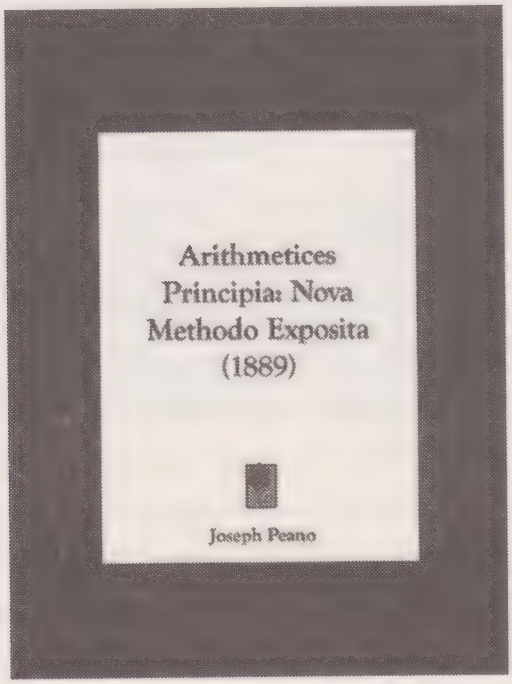
The axioms of arithmetic

In view of all the above, in order to define *arithmetic,* the first thing that needs to be done is to find some axioms for it. This search, towards the end of the 19th century, was of more relevance than what it might seem today, because while the dream of the first half of that century was to describe how the world was, during the second half the efforts focused on finding out precisely what natural numbers were. Other types of numbers, such as negative numbers and fractions, could easily be formed from them: so, -1 comes from adding a minus sign to the natural number 1, and is necessary when we have to distinguish between two directions, such as on thermometers or bank statements. At the same time, $2/3$ is formed by dividing 2 by 3, and responds to another necessity – splitting things

up when the result is not exact. But, what would have to be used to explain numbers that were not built from others?

Here we can find a variety of solutions: Georg Cantor (1845-1918) suggested defining a natural number as one that measures how many elements a set has, but as we shall see in the next chapter, the remedy was worse than the disease. That would no doubt have been a joy to his bitter enemy, Leopold Kronecker (1823-1891), for whom the question of how to describe the natural numbers was settled by the remark: "God created the natural numbers. All else is the work of man." It's a good job that Giuseppe (or Joseph) Peano (1858-1932) was not so shy, because if he were perhaps he would never have suggested a new point of view, which he outlined for the first time in 1889 in a book entitled *Arithmetices Principia: Nova Methodo Exposita*, that is, *The Principles of Arithmetic, Presented by a New Method*. Until that time, reasoned Peano, attempts had been made to first define the natural numbers and afterwards to find some axioms to prove theorems about them. But, why not do it the other way round? If we start with a list of axioms, then the numbers can be defined as those objects that verify them, even though besides *our* numbers, maybe we'll come across *others*...

This ingenious turnaround allowed Peano to erect a cathedral of arithmetic based on only five axioms, of which the fifth, known as the 'principle of



Book cover of Arithmetices Principia, Nova Methodo Exposita, *by Peano.*

induction', would once again be the one that was a little more difficult. The basic ingredients that intervene in arithmetic are a distinguished element, zero, and an operation, that to every natural number another is assigned, which we shall call its successor, or the following one. With this language, what the Italian mathematician proposed was to characterise the natural numbers as the objects that fulfil these properties:

a) Zero is a natural number.
b) Each natural number has a successor.
c) Zero is not the successor of any natural number.
d) Two different numbers have different successors.
e) If a set *A* contains zero and each time that it contains a natural number it also contains the following one, then *A* contains all the natural numbers.

The first theorem that can be demonstrated based on Peano's axioms is the one that states that one is different from zero, though for that we have first to explain what we call 'one'. A detailed examination of the demonstration will give us an idea of how axioms and rules of inference are manipulated. As we said earlier, proof that one is different from zero must necessarily begin with an axiom, which in this case will be 'Each natural number has a successor' (1st). Next we can make use of an axiom or of a statement deriving from the previous ones by a rule of inference. We again choose an axiom, in this case, 'Zero is a natural number' (2nd). Now the *modus ponens* enables us to deduce from the first two affirmations, 'Every natural number has a successor' and 'Zero is a natural number', the proof's third enunciation: 'Zero has a successor' (3rd). To abbreviate, we will call it one, and we'll show it as 1. At this point we can follow on with a re-writing of the axiom c) with the equivalent form 'If a number is zero, then it is not the successor of any number' (4th), and use the enunciation (3rd), which we already demonstrated in the first part and which says that 'One is the successor of zero'. This time using the *modus tollens*, we have 'If a number is zero, then it is not the successor to any number. As one is the successor of zero, then one is not zero'. And this is our theorem: 'One is different from zero' (5th).

With the certainty that one and zero are different numbers, what we could ask ourselves now is if the objects that satisfy Peano's axioms concur with the intuition that the natural numbers are a series that never ends or, in other

words, the numbers are infinite. To do that it would be enough for us to know that every number is different from the previous ones. And that is where the axiom of induction plays a vital role, which allows theorems to be demonstrated on all the natural numbers without having to verify it for every single one of them. One way of understanding what this principle consists of is to imagine the numbers as a row of dominoes, from which a few are chosen, and which we will distinguish from the rest by knocking them over. With that image, the principle of induction confirms what the reader expects: if we knock over the first domino in the line, every time that one falls the next one falls too. When we push over the first one all the rest will fall.

Once we have proven that there exists a natural number different from zero, which is called one, the argument can be repeated to prove there is still another number different from zero and one. Effectively, 'Every natural number has a successor' (1st) and 'One is a natural number' (2nd), then by applying the *modus ponens* it can be deduced that 'One has a successor' (3rd), which we will call two. By axiom d), which we will copy on the fourth line of the demonstration, we know that 'Two different numbers have two different successors' (4th). Now then, our theorem states that 'Zero and one are different' (5th), so by using the *modus ponens* again, 'The successor of zero is different to the successor of one' (6th), but those numbers are none other than those which we named, for ease of use, one and two. On the other hand, two and zero are different, because two is the number following one, and zero is not the successor of any natural number.

If we start up the reasoning again, by changing one for two we will demonstrate that there exists another natural number which we can call three, and that it is different from those needed to define it, that is, from zero, one and two. In general, by repeating the process enough times, we can deduce that a number in particular, for example, 1,729, is different from its successor and all the previous ones. Thanks to the axiom of induction, we know that to prove that 'every natural number is different from the following one' it is enough for us to be able to prove that one is different from zero (in other words, the first domino falls) and that the same goes for any concrete number and for its successor (or, expressed in different terms, that on knocking over a domino, the following one also topples).

By now, those kind readers who have come with us up to here will be doubting whether such long-windedness is really necessary just to be certain of something so elementary as the fact that two natural numbers are different.

And they are quite right to do so. No parent, if they have any love at all for their children, would use such an explanation to demonstrate that having two sweets is not the same as having only one. However, logic does not deal with how we reason in daily life, but rather with how we should do it to be sure that the conclusion we reach is the true one. What we have done is to strip the terms 'zero' ,'number' and 'following' of all their intuitive meanings and reduce them simply to abstract concepts that relate to each other through axioms and the rules of deduction.

What can we ask of axioms?

This new conception of axioms and proofs transformed theories, from those that favoured just a few evident truths into democratic systems in which all statements have the same right to become axioms. But that is only *a priori*. It would, for instance, be very unwise to stand and not intervene while a baby was elected head of the government, and such an intervention would not mean that a democracy had become less democratic. In the same way, if axioms are not chosen with care, the theories that emerge from them will be sterile. Euclid had a clear idea of how to do it, but as soon as the compass of experience disappeared, it was necessary to find formal criteria on the validity of the axioms, such as consistency, recursiveness or completeness.

To explain what it means when a system of axioms is consistent, we shall allow ourselves to fantasise a little about technology of the future. There is nothing to stop us supposing that within the next hundred years some evil group of scientists will design an infallible missile that will reach any target and destroy it in a matter of seconds. Or we could well imagine that, after very costly research into new alloys, the good guys' army will have managed to build an aeroplane able to withstand any type of impact. Separately, neither of these two scenarios would look out of place at the start of a science-fiction movie, but what the moviegoers would never accept is that both hypotheses were true at the same time, because if it should occur to anyone even more wicked than scientists – to a logician, for example – to fire the missile at the plane, we would have the paradox of a perfect projectile up against an indestructible target.

In general, we say that a set of axioms is consistent if it does not create contradictions, that is, if both a statement and its negation cannot be deduced from it at the same time. So, the axioms 'A perfect missile exists' and 'An indestructible

IN AN INCONSISTENT SYSTEM, ANY STATEMENT IS A THEOREM

Suppose that we want to demonstrate that a statement *Q* is true. As the system is inconsistent, there will be a theorem *P* the negation of which *notP* will also be a theorem. That means that we can find demonstrations of *P* and of *notP*. As we said before when studying the laws of inference, the deduction "If *P* and *notP*, then *Q*" is valid, as its hypotheses are never verified at the same time. Now then, in inconsistent theories *P* and *notP* are theorems, then by joining the rule of deduction "If *P* and *notP* then *Q*" and the demonstrations of *P* and of *notP*, the *modus ponens* enables us to prove that *Q* is a theorem. In other words, and however incredible it might seem, in the worlds in which zero is equal to one and different from it at the same time, you are my father (even if you are a woman). *From a contradiction, everything follows...*

aeroplane exists' are not consistent, because from the first it follows that when the missile hits the plane it will destroy it, and from the second, that the plane will remain intact. The word 'consistent' is from the Latin *consistentem*, and means coherent, not contradictory. That is the minimum that can be demanded of axioms, because in theories that are not consistent any proposition is true, and speaking about everything comes to the same as speaking about nothing. The problem is that to be sure that a system of axioms is consistent, we often have to resort to theories that are more complex and the consistency of which raises more questions than answers.

To introduce the concept of completeness, we shall switch to the crime genre and use an example inspired by the Argentinian writer Guillermo Martínez. Let's imagine that a murder has been committed in a locked room, and that when the police arrive, beside the dead body they find two suspects. Each of them knows who did it. Each knows if it was or was not he himself who did it. However, unless there is a confession, the detectives will have to resort to fingerprints, traces of DNA or any other evidence enabling them to convict one of them. If this search is inconclusive, they will remain free, but the truth of what happened will always be there. Although the truth exists, the means are insufficient to reach it.

After a hard day's work, the police officers go out for a drink to relax. One of them has only just come to work at the police station and the others hardly know him. From what he tells them of his life they can deduce that he was born in Manchester, but that almost at once he was taken to live in Brighton because

his parents liked to be near the sea. However, with the data they have on him, his colleagues just cannot reach agreement on whether he is married or not, even though there is no doubt that only one answer is possible.

Both of these situations show that in many aspects of daily life, the truth does not coincide with what is provable. That is what logicians mean when they say that a system of axioms is not complete. The ideal would be for it to be possible for all true statements on certain objects to be provable just from a handful of axioms. But that rarely happens. Normally a theory contains statements that cannot be demonstrated nor refuted, which we shall call 'undecidables'. Refuting a statement is understood as proving its negation. For instance, to refute the statement 'All swans are white', which we mentioned earlier, would mean proving that 'There exists a swan that is not white'. Complete theories are those that do not contain undecidable statements, or, what comes to the same thing, they are the systems of axioms in which any proposition can either be directly proven, or first negated then proven. Attentive readers will have noticed that, in this second definition of completeness, the hazy concept of truth has been replaced by that of proof. That's how solutions were found for some of the paradoxes that had occupied philosophers since antiquity.

With most mathematical theories the same thing happens as with the jury who cannot make up their minds if the suspects are guilty or innocent. So it will come as a surprise that we now explain that there is always a way to choose axioms so that a theory becomes complete. It consists of including all the true statements. With this process of axiomatisation, proofs only have one line, as what it being proven is already an axiom. If paradise for logicians is theories that are complete, why not do it like that? The provable would coincide with truth, and the demonstrations are as short as can be. However, the set of all the true propositions is too large for us to be able to take them as axioms. We are not so much interested in the length of the proofs as in being able to check that they are correct by means of some automatic procedure. In a proof, each line is an axiom or is deduced from the previous ones by applying rules of inference. In order to find out if a list of statements proves a theorem it is essential for us to be able to verify if a proposition is an axiom. However, if we choose too many the time needed is infinite.

We say that an axiomatic system is recursive when that does not happen, in other words, when it is possible to verify, in a finite number of steps, if an affirmation is or is not an axiom. Recursiveness puts a brake on the greed of

logic, which wants to prove more and more theorems, because it prevents it from adding all the axioms necessary to complete a theory. Of course, geometry and arithmetic are recursive theories, as are generally all those theories in which there is only a finite number of axioms. However, there are also recursive systems with infinite axioms. But that does not matter, because the most important thing about recursive systems is not how many axioms they have, but that the validity of any proof based on them should be able to be verified in a finite number of operations.

A RECURSIVE SYSTEM WITH AN INFINITE NUMBER OF AXIOMS

One of the ways to obtain a recursive system with infinite axioms is to lay out one of Peano's axioms in infinite affirmations. To some extent, 'Zero is not the successor of any number' is simply a shortened way of saying 'Zero is not the successor of zero', 'Zero is not the successor of one', 'Zero is not the successor of two', and so on indefinitely. Let's suppose now that we want to verify if a statement is one of these axioms. It will only be in the list if it begins with 'Zero is not the successor of...', and if what follows is a number. By remembering that 'one' really means 'the successor of zero'; 'two', 'the successor of the successor of zero', etc., the only thing that needs to be done is to count how many times the word 'successor' appears in our statement. Therefore, this axiomatic system is recursive as well as infinite.

Let's recap. The axiomatic method appeared around 300 BC with the *Elements*. Euclid considered axioms to be evident truths, in accordance with our experience of physical things, but the discovery of other geometries different from his, halfway through the 19th century, shot down this realist conception. Since then, axioms are only statements that are chosen for the sake of convenience as the basis of mathematical research. When we apply certain rules of deduction to axioms, such as *modus ponens* or *modus tollens*, we get new true statements that we mathematicians call theorems. Precisely the truth of those theorems lies in the proofs, which are finite chains of statements, of which the first one is an axiom and the following ones are either axioms or they are deduced from the previous ones by rules of inference. A theory is a set of axioms, of rules of deduction, and of all the theorems that can be proven with those components.

Logic is the branch of maths that deals with the study of theories, disregarding what they say. Faced with an axiomatic system, a logician is not interested

in its contents but in three formal questions: consistency, recursiveness and completeness. The first of these ensures that no contradictions will be produced within our theory and is the minimum guarantee necessary to be able to erect the mathematical construction on solid foundations. Recursion, on the other hand, makes sure that there are not too many axioms, because if there were, we would be running the risk of not being able to determine if a demonstration is or is not correct. Finally, the completeness of a theory tells us when its axioms are sufficient to deduce all the true affirmations they refer to, or expressed in another way and without making use of the concept of truth, it tells us when any proposition can be demonstrated or refuted.

In the next chapter we will study a series of paradoxes which, at the end of the 19th century, caused the foundations of over two thousand years of mathematics to tremble. Luckily, several solutions were soon put forward, all of them stipulating that it was not enough for axioms to appear to be consistent. They had to be proved to be so! We will look at this *formalist program* in Chapter 3. And equipped with all that baggage, we shall then be able to wonder at one of the most beautiful results in logic: Gödel's incompleteness theorem, which establishes a balance between the notions of consistency, recursiveness and completeness.

Chapter 2

The Paradoxes

Paradox is the passion of thought.
Søren Kierkegaard

Although Bertrand's parents' will tried to ensure that the youngest member of the Russell family would be brought up according to the same principles that they had fought for in Victorian Britain, his surviving and strong-willed paternal grandmother was not willing to let the monsters of atheism pervade the mind of this intelligent little boy. Within the strict atmosphere of the family home, one governess after another was to lead Bertrand along the paths of religion and languages, though it must be said that they did enable the young aristocrat to master French, German and Italian, which years later would allow him to travel the world without problems. But in those far-off days of his childhood, Bertrand only thought about the exotic Greek characters, as he found them so appropriate for expressing his unhappy reflections on the life he had been fated to live.

This melancholy did not leave him when he was sent to the academy at Old Southgate to prepare for the exams to get into Cambridge University. He had imagined, poor boy, that contact with other boys of his age would help him to shrug off the weight of his sadness! In his imagination he had seen an idyllic ambience in which he would be able to read the great English poets and discuss them with other pupils, or to be up till dawn talking over the philosophical issues that intrigued him. He found, however, a bunch of thuggish youngsters who only wanted to get drunk and chase after women, and who didn't miss the slightest opportunity to make fun of the shy boy who had always been pampered. Like a despairing romantic hero, many an evening Bertrand walked over the fields on his way home to New Southgate watching the Sun setting... and contemplating suicide.

If he did not kill himself it was not because of a lack of courage, but because at the age of 11 his brother Frank had opened the doors for him to a paradise in which he could take refuge, where there was still so much for him to explore. Young

Bertrand Russell in 1893, at the age of 21, graduating as a Bachelor of Mathematics at Trinity College, Cambridge.

Bertrand's entry into the garden of Euclid's *Elements* was as dazzling as a first love, a garden to which he ran whenever the hostility of the world around him became unbearable. But that happiness was clouded by the idea that even though the Greek wise man proved it all – according to what his brother had told him – the first thing demanded of the reader by those pages that Frank spelt out to him every evening was to make an act of faith just like those his grandmother demanded: 'A point is that which has no parts'. But what if it did have? 'A straight line can be drawn passing through two points'. What if it couldn't? Reluctantly, Bertrand had to listen to the advice given him by his brother, who strove in vain to make him see that if he did not believe the axioms, it was hardly worth continuing with the studies.

A good while later, twelve years after the day he arrived at Old Southgate Academy, Bertrand was again as shut off from those around him as when he used to go for walks thinking about suicide. In the meantime, many things had happened to him. He had graduated in mathematics and philosophy at Cambridge, where a secret society made up of the best students, calling themselves the Apostles, had at last provided him with the thousands of hours of conversation that years before he

thought he would find at school. He had travelled, published his first books – on German social-democracy and on the basics of geometry – and he had even had time left over to marry Alys Pearsall, the daughter of an American Quaker family. Despite everything, his main occupation was still mathematics and his goal to reduce the axioms of geometry to the laws of logic. No more would things have to be believed simply for the sake of it.

When trying to deduce the whole of mathematics from logic, Bertrand had come across a contradiction that at first sight appeared to be one of those riddles of the type 'Can a man marry his widow's sister?' For such a puzzle, it was enough to examine the meaning of each term to spot the trick (he's dead, so no he can't). However, the solution to the contradiction that worried Bertrand Russell would demand a lot more effort. Day after day for two summers he sat before a blank sheet of paper. Russell had seen the long mornings and the slow afternoons go by with the sheet still blank before he finally came to the conclusion that the set of all the sets that are not members of themselves does not exist.

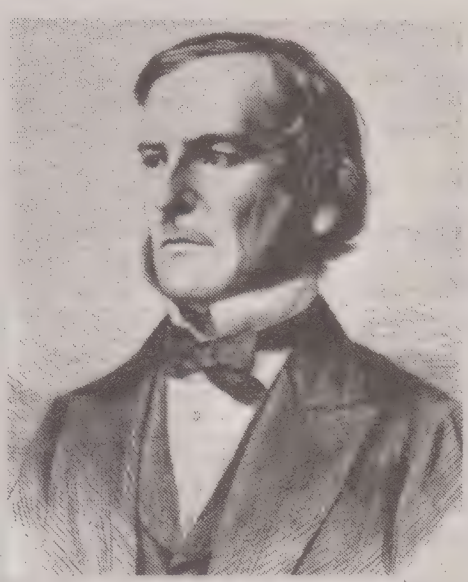
Set theory

To understand the contents of that paradox, which put an end to Bertrand Russell's 'glad confident mornings' we first need a few basics of set theory. In the previous chapter we attempted to show how the structure of the axiomatic method had already appeared in the *Elements*. However, Euclid's axioms were self-evident truths and not principles admitted for the sake of convenience. Over time, the Euclidean language had proved to be insufficient for transmitting the new mathematical ideas. To demonstrate the profound theorems of the 19th century just by words and diagrams had been as difficult as it would be for us today to translate the instruction manual of a smart phone into Latin.

Little by little, mathematical writing was becoming more symbolic. Not only did it have an extensive notation system for the series, the derivatives and the integrals but, thanks to the research of the English mathematician George Boole (1815-1864), statements in logic could also be written in the form of an equation. Geometry studied the forms of space; arithmetic, the numbers; analysis, the necessary tools for formalising the laws of physics; and algebra focused on equations. Would it be possible to find a common language for all these disciplines that would demonstrate the unity of mathematics?

THE ALGEBRA OF BOOLE

George Boole was one of the first to realise the analogy that exists between the connectors 'or' and 'and' in logic and the operations of adding and multiplying in algebra. He also introduced the symbols 0 (false) and 1 (true) for the two possible values of true. Before looking at an example, let's remember that when two numbers are multiplied the result is zero only if one of them is zero. Let's suppose that we want to translate the proposition 'All men are mortal' into algebra. Boole proposed using p to denote the value of truth of the statement 'to be a man', and q, to that of 'to be mortal'. With this ability, all the contents of the sentence could be reduced to the equation $p \cdot (1-q)=0$. Effectively, if someone is a man, then p takes the value of truth 1 (true). The equation tells us that the product of the numbers p and $1-q$ is zero. As p is different from zero, the only possibility left is that $1-q$ has the value 0. But that means that q is equal to 1 (true), in other words, man is mortal.



George Boole, one of the pioneers of computational algebra.

By reflecting on a problem that at first had nothing to do with this question, which is a more philosophical issue than mathematical, Georg Cantor believed he had found the answer in set theory between 1878 and 1884. Basically, a set is nothing more than a collection of objects. We could talk of a set of pens and pencils, a set of golf clubs or a tea set. These collections can be defined by extension, that is to say, by means of the list of elements that form them, or by comprehension, by indicating the property that is common to all its members. Hence the set of the natural numbers (remember, they are the numbers we use for counting) is none other than $\mathbb{N}=\{0, 1, 2, 3...\}$; this is an example of definition by extension. If we wanted to study just the even numbers, we would write $2\mathbb{N}=\{0, 2, 4, 6...\}$, or alternatively $2\mathbb{N}=\{n\in\mathbb{N}\,|\,n$ is divisible by $2\}$, where the symbol $\in$ means 'belongs', and the vertical bar $|$ means 'such that'. In this case it is definition by comprehension, as here we are considering the *subset* of the natural numbers with the property of being divisible by two.

As soon as he started his research, Cantor realised that his new theory dealt with two objects of a radically different nature at the same time – both finite sets and infinite sets. In fact, the problem of the calculation of the number of elements of a set, what we mathematicians call its *cardinality*, had very different solutions depending on whether the set was finite or infinite. Let's begin with a very simple situation. Suppose we want to know whether two finite sets have the same cardinality, for example, if there are as many letters in the word 'insolence' as colours in the rainbow. The obvious procedure is to count the elements that make up each set and then compare the results. As I-N-S-O-L-E-N-C-E has nine letters, while red, orange, yellow, green, blue, indigo and violet come to seven colours, we say that the two sets have a different cardinality. But, what would happen if we tried to apply the same method to two infinite sets? The only thing we could conclude in this case is that they are infinite, so either we accept that from the point of view of the cardinality all the infinite sets are equal, and the game is over, or we find ourselves forced to modify the method.

Coming back to finite sets, let's see what happens if instead of dealing with the two collections separately we work along simultaneously extracting an element from each of them: we would start with the letter *I* and the colour red, we follow with the *N* and with the orange, until we arrive at *N*, which corresponds to the colour violet. At this point, one of the two sets has ended, but the other still has two elements left, the letters *C* and *E*, so we can conclude that its cardinality is greater. What we have tried to do here – without success – in mathematics is called establishing a bijection between two sets. It means to assign to each element of a set *X* an element from another set *Y, one to one*. We do it in such a way that the following requirements are met:

1. There are not two elements of *X* to which the same element of *Y* corresponds.
2. All the elements of *Y* are paired with some element of *X*.

With this terminology, two sets have the same cardinality when there is bijection between them. It is easy to show that between two finite sets with a different number of elements, a bijection cannot be established, as several elements of *X* would necessarily end up going to the same term of *Y*, or some element of *Y* would remain unpaired.

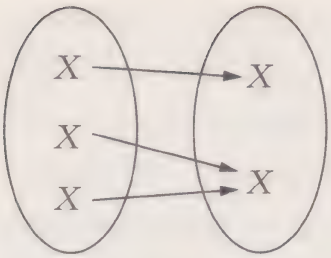


Fig. 1

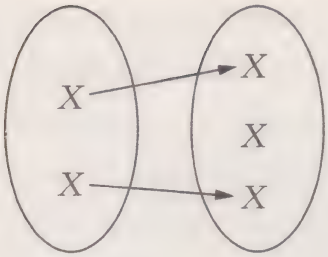


Fig. 2

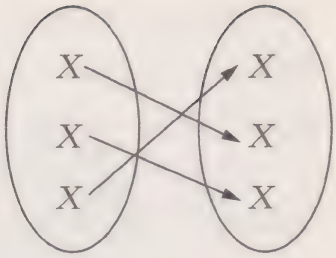


Fig. 3

Three examples of correspondences between finite sets, of which only the last one (fig. 3) is a bijection. In fig. 1 two elements end up at the same point and in fig. 2 there is an element in the destination set that has not been paired.

The advantage of this new approach is that we can extend it to the infinite sets. We will say, therefore, that two infinite sets have the same cardinality when there is bijection between them. The first consequence will perhaps surprise the reader: there is the same quantity of even numbers as even numbers and odd numbers all together. How can that be possible? In accordance with our definition, to demonstrate this result, which is counter-intuitive, it would be enough to define a bijection between the natural numbers and the even numbers. We suggest the following, we will make these correspondences: 0, 0; to 1, 2; to 2, 4, and so on, that is, to each number n, twice that number. We can at once see that with this method different numbers are always sent to different numbers, and that any even number sooner or later appears paired with its half. As properties 1 and 2 are verified, there is the same amount of even numbers as there are of numbers!

Allow us to reformulate the result further: 'In a hotel with infinite rooms there is always room for more guests, even if the hotel is full.' In reality, in those hotels that have a finite number of rooms, when all of them are taken the best situation we can hope for is that the receptionist will give us the address of another hotel where we can stay. This never happens in the infinite hotels: there is the same number of rooms as even-numbered rooms. We could use the bijection that we have created to switch the guest in the first room to the second; the one in the second to the fourth, etc., and therefore free up all the odd ones. We have not only made room for the traveller needing a room but also for an infinite number of other travellers who will arrive next. Hoteliers would do well to take note...

Far from being a mere curiosity of even numbers, the existence of such hotels that never fill up is the essential characteristic of infinite sets, as Richard Dedekind

pointed out in his article *What are Numbers and What Should They Be?* published in 1888. A set is infinite if it can be put into bijection with a part of itself that does not contain all the elements. It is obvious that this will not happen if we start off with a finite set as it lacks some elements it will not be possible to put it into bijection with the total. (As we said above, two finite sets, of m and of n elements respectively, can only be put in bijection if $m = n$). However, natural numbers are infinite because a part strictly contained in them, the even numbers, has the same cardinality as the whole set. The new definition, therefore, coincides with the reasoning based on Peano's axioms, which in the previous chapter enabled us to prove that natural numbers were infinite. In fact, it is the smallest infinite set that we can imagine. That's why all the sets in bijection with natural numbers are known by a special name: they are the countable sets, and their cardinality is denoted by the first letter of the Hebrew alphabet, the *aleph*, with a subindex that indicates that it is the smallest infinite cardinality: $\aleph_0$.

What does it mean for a set to be countable? As we have seen, it is just an abbreviated way to say that X can be put into bijection with the natural numbers. So, for each natural number n there will be a corresponding element of the set which we shall call x_n, such that, on the one hand, if n and m are different, then x_n and x_m are different, and, on the other hand, all the elements of X can be written as x_n for an n. When we used to go on a school trip, the teacher used to give us all a number to make sure we all came back. Before getting on the bus, we used to shout out the series of numbers: number one! number two! number three! We all had a number, and none of them were repeated. Countable sets can also be made to shout out their position in the list: at the call of 'number one!' it will answer x_1, and when we call out 'number two!' x_2 will appear. Countable sets are those that can be put in a row. We saw that even numbers are countable because there is an evident way to arrange them: 0, 2, 4, 6, 8, 10... The same thing happens with positives and negatives, as beginning with 0 we can skip off to each side: 0, 1, –1, 2, –2...

But can't the elements of any set be placed in a row? In that case all sets would be countable, and we would not have got any further than where we were at the beginning, when we made do with just counting. The reader has no need to worry on this point, as one of Georg Cantor's great discoveries was the existence of uncountable sets. Let's look at the simplest example: the set formed by the infinite sequences of 0*s* and of 1*s*, in other words, by objects of the form 0100100010... or

1100101001... We'll show that in the event that we are dealing with an countable set, we'll immediately come across a contradiction. If it were in fact an countable set, we would be able to write all its elements in a list like this:

First element	→	$\boxed{a_0}$	a_1	a_2	a_3	...	...
Second element	→	b_0	$\boxed{b_1}$	b_2	b_3	...	...
Third element	→	c_0	c_1	$\boxed{c_2}$	c_3	...	...
...							

Remember that in the list the symbols a_n, b_n and c_n only take the values 0 and 1. We are going to construct an element which, despite belonging to the set of infinite sequences of 0*s* and of 1*s*, does not appear in our list. To do so, we'll take note of the terms in the diagonal we have highlighted in boxes. Let's take a look at a_0: if the value is 0, we begin our sequence with 1, and if the value is 1, with 0; this determines the first term for us. Let's continue with b_1: if the value is 0, then the second term in our selection will be 1. If by contrast it is equal to 1, then we will write 0. For all, to determine the *n*th degree term of our sequence we look at the corresponding element in the diagonal and write the opposite symbol. Thus we obtain a sequence entirely comprised of 0*s* and 1*s*, and which, therefore, forms part of the set. For example, if the beginning of the list were

First element	→	$\boxed{0}$	1	0	0	...	...
Second element	→	1	$\boxed{1}$	0	0	...	...
Third element	→	0	0	$\boxed{1}$	1	...	...
...							

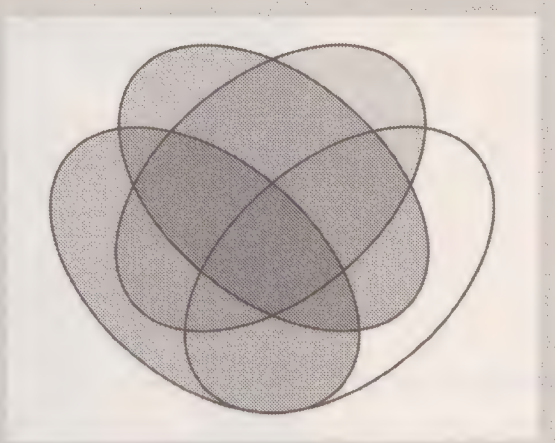
then the object that we are constructing would begin with 100... As it consists of modifying the elements of the diagonal, this method of constructing a sequence of 0*s* and of 1*s* based on the hypothetical list is called diagonal argument. What we want to see is that the sequence that is deduced from the diagonal argument, despite forming part of the set, does not appear in any position in the hypothetical list that

contained all its elements. Effectively, it cannot be the first sequence, because the first term is different; nor the second, as we have varied the second term; nor the third, nor the fourth. Each sequence in the list has at least one element different from the one we have constructed, specifically the one that appears in the diagonal. We had supposed that the set of the sequences of 0*s* and of 1*s* was countable, in other words, that all its elements could be placed in a row, but we have reached a contradiction. That proves that it is uncountable!

We wanted to devote a few pages to explaining details of the basic concepts of set theory, and it was not only so that we will be able to formulate Russell's paradox better in the next section. The proof that the set of sequences of 0*s* and of 1*s* is not countable, which at this point the reader might be thinking is just a display of virtuosity, will allow us in Chapter 5 to demonstrate that there are tasks that even computers cannot carry out. And we hope that on the way have realised how bizarre the world of infinite sets is.

SET THEORY IN SCHOOLS

During the 1970s a group of extremist followers of the French Bourbaki Society, the majority of whom were not mathematicians, tried to get set theory taught at primary school level. If any reader had to endure this pedagogical overkill then for sure he or she will remember that the natural numbers were explained as the cardinalities of the finite sets: 0 is the cardinality of the empty set, and to add 2+3 it's enough to join a set of two and another of 3 elements; it doesn't matter that the result is called 5, the important thing is that $2+3=3+2$, as it doesn't matter in what order we mix the sets. According to Pierre Cartier, then secretary of the Bourbaki Society, the natural outcome of that educational policy was that children came home from school sobbing: "Mummy, I don't want to be a set".

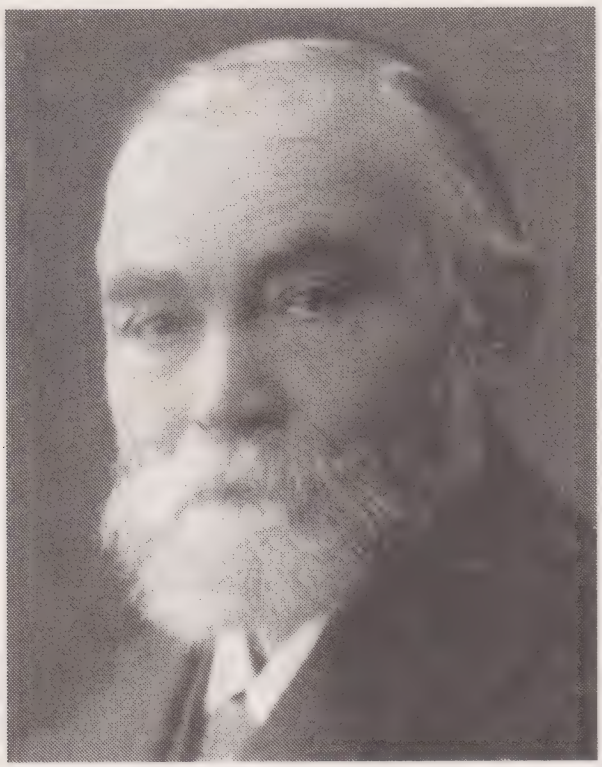


Venn diagrams are the most common way to represent sets.

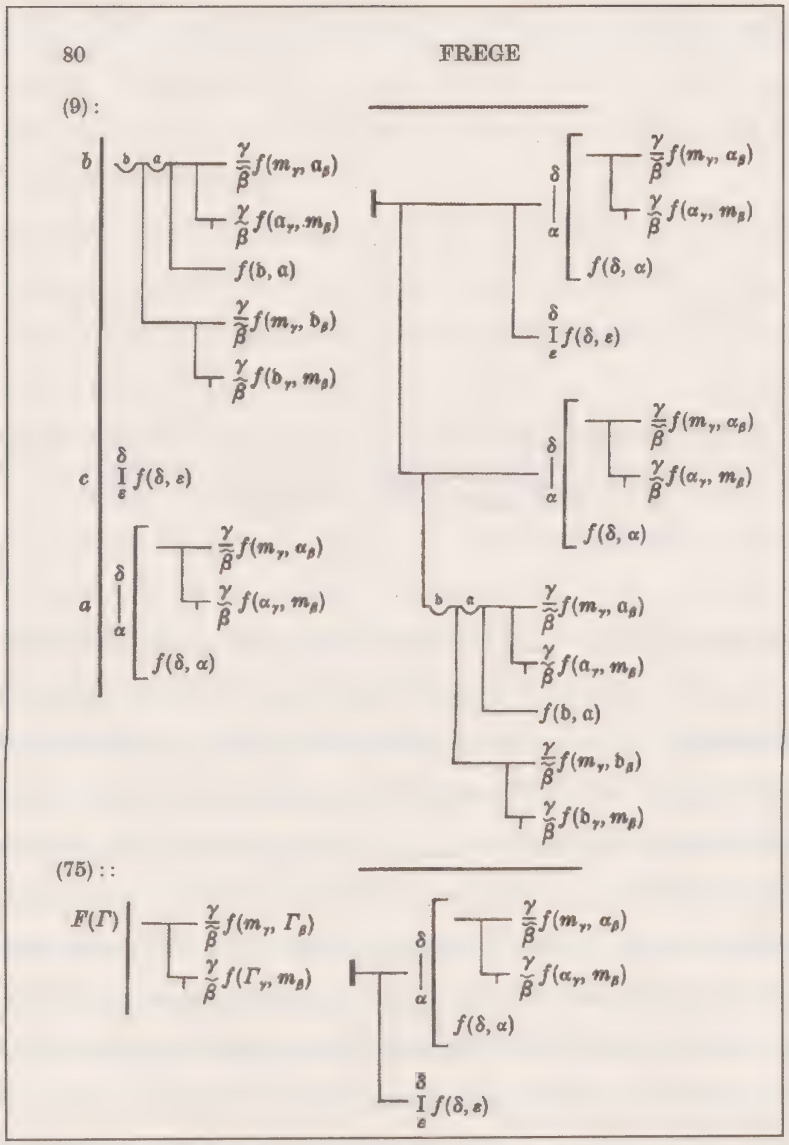
The Russell paradox

Bertrand Russell had learnt about set theory in 1896. At first it took him a while to accept it, because the author of the book about it was not a supporter of Cantor's ideas and doubted their rigour. Instead they were presented as theology in disguise, which was precisely the opposite of what interested Russell at that time. Later, however, he realised that many of the accusations were unfounded and he included the German mathematician's point of view in the final edition of *The Principles of Mathematics*, published in May 1903. While reading the latest academic literature, Russell discovered the work of Gottlob Frege, who had foreseen many of Russell's discoveries 20 years earlier. It was not always easy to spot that they were the same ideas, because Frege's complicated symbolism, similar to a contemporary music score, was nothing like the clear notation that Russell had learnt from Peano.

By closely studying *Ideography (Begriffsschrift)*, the book in which Frege had first set down his research, Russell had begun to reflect on the set of all the sets that are not members of themselves. The set of all the cats is certainly not a cat, but the set of all things imaginable is, in turn, something imaginable. We say that these sets 'belong to themselves' or that 'they are members of themselves'.



Gottlob Frege is the father of mathematical logic.



A page from Ideography *by Gottlob Frege.*

As the reader will agree that this is a somewhat confusing property, we would like to eliminate all sets of this type at a stroke. To do so, we shall give the name R (R for Russell) to the set of all the sets that are not members of themselves: in R is to be found the set of cats, as is the set of tables, and in general, all the collections that have the good taste not to belong to themselves, so we will be safe as long as we don't cross the boundary separating R from the other sets.

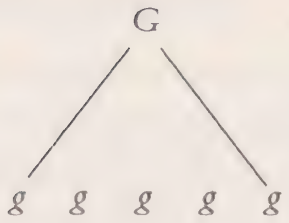


Fig. 1

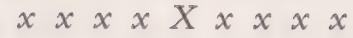


Fig. 2

The difference between the set of all the cats G, which is not a cat (fig. 1), and the set of all the things imaginable X, which is one of the imaginable members in the set (fig. 2). (Source: Umberto Eco, The Infinity of Lists*)*

The paradox arises when we ask ourselves on which side of the border *R* itself is, because any answer implies the contrary. Indeed, let's suppose that the set *R* is a member of itself; then *R* satisfies the property that we wanted to eliminate, so it cannot belong to the set of all the sets that are not members of themselves. But, what set is that? *R* again! Therefore, if *R* belongs to itself, then *R* does not belong to itself. Up to here there is no problem, because it could occur that *R* were not a member of itself and that, based on this hypothesis, no contradiction at all would be found. However, let's see what happens when we suppose that *R* does not belong to itself. In that case, *R* would fulfil the property that defines the set of all the sets that do not belong to themselves, so *R* would be included in it. That is to say, if *R* is not a member of itself, then *R* is a member of itself. Together, both conclusions breach a basic principle going back to the philosopher Parmenides, who had shown in his didactic poem *On Nature* that there are no intermediate channels between being and not being. In its mathematical version, the principle states that an element belongs or does not belong to a set. As any third possibility is excluded, we mathematicians refer to it as the axiom of the excluded third.

To explain his paradox in simpler terms, Russell imagined a little village whose laws obliged the barber to shave all the people who did not shave themselves and no one else. We have replaced the property of 'being a member of oneself' with that of 'shaving', so the barber now takes the role of set *R*. The paradox stems from the question: who shaves the barber? Because if he shaved himself he would then belong to the group of people that the law bans him from shaving. But if he didn't do so, then he would be under the legal obligation to shave

himself. Whatever he does, the barber will end up in prison, where perhaps a logician might try to convince him that spending a few years in jail is preferable to discovering a contradiction which would throw doubt on the validity of two thousand years of mathematics.

Another version of the paradox replaces the barber with a serious-looking librarian who has been given the task of putting some order into a library that is so big that it needs a catalogue to contain all the catalogues. Someone suggests that a good criterion to use would be to separate the catalogues that don't mention themselves from those that do. As he finds the idea of this fight against bibliographic narcissism amusing, the librarian immediately gets to work. After working day and night for several years, all the shelves have been examined one by one, and the only thing left to do is to decide where to list the volume into which he has put so much time and effort. If it lists itself, then he cannot include it in the catalogue of all the catalogues that do not list themselves. If, on the other hand, this catalogue was not to be one of its own entries, then it should form part of the catalogue of all the catalogues that do not list themselves (and thus be on the list). If it belongs, it doesn't belong, if it doesn't belong, then it belongs. Only then did the librarian realise that his work had been in vain, as the criterion would never allow a complete classification to be compiled.

Shortly after discovering the paradox, Russell wrote a letter to Gottlob Frege, who then corrected the proofs in the second part of his magnum opus, *The Foundations of Arithmetic* (published in 1884). In the work he included an axiom thanks to which it was possible to form the set of all the objects that satisfied a property P, but Russell had discovered that axiom, applied to P = 'to be member of itself', led to a contradiction, as the set R of all the sets which do not belong to themselves breaches the axiom of the excluded third. Filled with dismay at the news, but without losing sight of his characteristic rigour, Frege added an epigraph in which he confessed that "hardly anything more unfortunate can befall a scientific writer than to have one of the foundations of his edifice shaken after the work is finished". Later, he proposed modifying his axiom, but the alternative was not consistent with the rest of the system, and so several years had to go by before Russell's paradox was solved.

Between 1906 and 1908, Russell found a simple solution to his paradox, with which he would lay the foundations for the type theory. He had earlier been troubled by the ontological problem created by descriptions like 'the greatest natural number'

RUSSELL ON FREGE

In a letter sent to the historical logician Jean van Heijenoort dated 23 November, 1962, Bertrand Russell spoke of Gottlob Frege in these terms:

> "As I think about acts of integrity and grace, I realise that there is not anything in my knowledge to compare to Frege's dedication to truth. His entire life's work was on the verge of completion, much of his work had been ignored to the benefit of men infinitely less capable, his second volume was about to be published, and upon finding that his fundamental assumption was in error, he responded with intellectual pleasure clearly submerging any feelings of personal disappointment. It was almost superhuman and a telling indication of that of which men are capable if their dedication is to creative work and knowledge instead of cruder efforts to dominate and be known".

or 'the present king of France', which, while being grammatically correct, did not, however, refer to anything. The case of the 'set of all the sets that don't belong to themselves' is even worse. It is not that it doesn't exist, but that not even the description that defines it is valid. It would be like speaking of 'the France of the current king' or 'greatest the number natural'.

In the simplest version of Russell's theory, every mathematical object can be assigned a number depending on its complexity. The elements are of type 0, the sets of elements are of type 1, the sets of sets of elements are type 2, and so on. For instance, if we think of natural numbers, number 8 is type 0, the set E of all the even numbers, and the set I of all the odd numbers are type 1, and when we take set $\{E, I\}$ we have moved to type 2, because its elements are now of type 1. Once a type has been assigned to each object, there is an unbreakable rule: we can only affirm the membership of an object of type n to another of type $n+1$. The expression 'number 8 is even' is correct because 8 is of type 0, and set E is type 1. However, it makes no sense to ask ourselves if the set E of the even numbers is or is not an even number, as this is a membership relation between objects of the same type. Now then, this was precisely what was described when we spoke about the set of all the sets that are not members of themselves. In the language of logic, 'to be a member of itself' is conceptually incorrect, and in this way the

Ernst Zermelo, the first axiomatiser of set theory.

paradox disappears. It is true that given a property *P*, the set of the objects can be considered to fulfil it, but the least that can be asked of *P* is that it be well defined.

While Russell was publishing *Mathematical Logic based on the Theory of Types* in the *American Journal of Mathematics*, Ernst Zermelo (1871-1953) was proposing a new solution to the paradox which was less conceptual than Russell's but infinitely more practical for 'mathematical workers'. Nowadays, we realise that one of the greatest difficulties when starting off a theory is to define the object of study. Everywhere we hear people talk of the science of information, but what is information? Some would define biologists as those who study life, but what is life? Those are the sort of questions that Zermelo asked himself about set theory. According to Cantor's intuitive idea, sets were no more than collections of things that fulfilled a certain property, but that allowed the set of all the sets that were not members of themselves to be constructed. Without a precise definition, there was no hope of getting very far. What Zermelo did was to replace the ingenuous notion of set with a list of axioms, among which was one that prevented Russell's paradox from being formed. From then on the sets would be the objects that verify the list of axioms.

The liar paradox

Though we started this chapter with an analysis of the Russell paradox we would not want anyone to be deceived into thinking that paradoxes are a product of our contemporary times. The etymology of the term itself, *para-dox*, which means that which is outside common opinion, points unmistakably to its Greek roots. In a wider sense, a paradox is the absurd reached from reasoning that starts off with plausible hypotheses and continues with logical deductions that appear to be valid. So when Russell concerned himself with the set of all the sets that are not members of themselves, behind him stretched a long literary and philosophical tradition. However, until the end of the 19th century it did not seem possible that paradoxes would cross over the borders of the humanities to attack the realm of the purest reason. Philosophers had made use of paradoxes to deny the illusion of the senses, and the poets as the only vehicle to tell the truth about love; but mathematicians feared them as a 'Pandora's box', which, once opened, could destroy everything in an instant. That's why the discovery of the contradictions to which set theory was leading, at a time when the work of Cantor was beginning to be accepted as a universal language for mathematics, produced a fundamental crisis from which the science would take years to recover.

One of the oldest paradoxes is the one about Achilles and the tortoise, with which the pre-Socratic philosopher Zeno of Elea, a follower of Parmenides, wanted to show that movement did not exist and in this way attack the defenders of an atomistic conception of space and time. The advantage that Achilles gives the tortoise so that they will both compete in equal conditions – explained the Greek – represents an insuperable breach, as by the time the athlete has run to the initial position of the tortoise, the animal will have moved a little. When Achilles reaches the tortoise's new position, he will still not be able to catch him, because the tortoise will have moved a little. Of the space that separates them there will always be a fraction left, however small it may be, which will prevent the flight-footed man from winning.

In another equivalent formulation, Zeno states that an arrow will never reach its target, because when it has covered the first half of the distance to the target, it will have to cover the other half; when it has covered the half of this distance, it will still have a quarter left; when it has covered half of this quarter, it will still have an eighth left, and so on into infinity. However, in practice, Achilles beats the tortoise and the arrow hits the target.

Perhaps the most intriguing classical paradoxes are the antinomies, statements that are true and false at the same time. One of the most outstanding is the liar paradox, which is generally attributed to Epimenides of Crete, although it's possible that the philosopher, who was said to have fallen asleep and slumbered for 57 years in a cave blessed by Zeus, was unaware that he was saying it. In a verse in his poem *Cretica*, Epimenides attacks the "Cretans, always liars" who did not believe in the immortality of Zeus. But, because he was Cretan too, his statement referred to himself became 'I always lie'.

Let's suppose that Epimenides is actually lying; then what he says cannot be true, and, as he says he is lying, he must be telling the truth. If, on the other hand, Epimenides is telling the truth, then what he is saying must be true, and as he says that he is lying, he must be lying. According to legend, the poet Philitas of Cos died of exhaustion trying to find an answer to the paradox.

In actual fact, 'I always lie' is not a paradox in the strict sense, as it's negation is not 'I always tell the truth', as we insinuated in the preceding reasoning, but 'I don't always lie', or alternatively 'I sometimes tell the truth'. However, if the words "This sentence is false" are put into Epimenides mouth, the result is an authentic paradox. Indeed, let's suppose that the sentence is true: then what he says must be a fact, so it's false. But if the sentence is false, as it is simply what it says of itself, it must necessarily be true. If it is true, it is false; if it is false, it is true. This breaches the *principle of bivalence*,

THE ISLAND OF KNIGHTS AND KNAVES

A logician arrives at an island whose inhabitants are split into two types: the knights always tell the truth, and the knaves always lie. On meeting the inhabitants *A*, *B* and *C*, the logician asks *A* if he is a knight or a knave, but the reply is so unclear that he finds himself forced to ask *B*: "What did *A* say?" *B* replies: "*A* said that he is a knave." But just at that instant *C* interrupts and warns the logician: "Don't believe *B*, he's lying!"

With these two pieces of information, the logician is able to identify *B* and *C*. In effect, according to *B*, inhabitant *A* said "I am a knave," which is another version of the liar paradox: "I always lie" – and knaves always say they are knights. Therefore, the only non-contradictory explanation is that *B* lied when transmitting *A*'s information, so *B* is a knave. So *C* was telling the truth when he warned the logician, from which it follows that *C* is a knight. Unless we carry on asking, we won't have enough information to determine what *A* is.

according to which a sentence is either true or false, and the *principle of contradiction*, which states that both situations cannot occur at the same time.

Every epoch has reinterpreted the liar paradox its own way. Cervantes, for example, remakes it in Chapter 51 of the second part of *Don Quixote* (On the progress of Sancho Panza's government and other such entertaining matters) to present it as an example of the difficult decisions that Sancho Panza will have to make in charge of the island of Barataria. Previously, in Chapter 18, Don

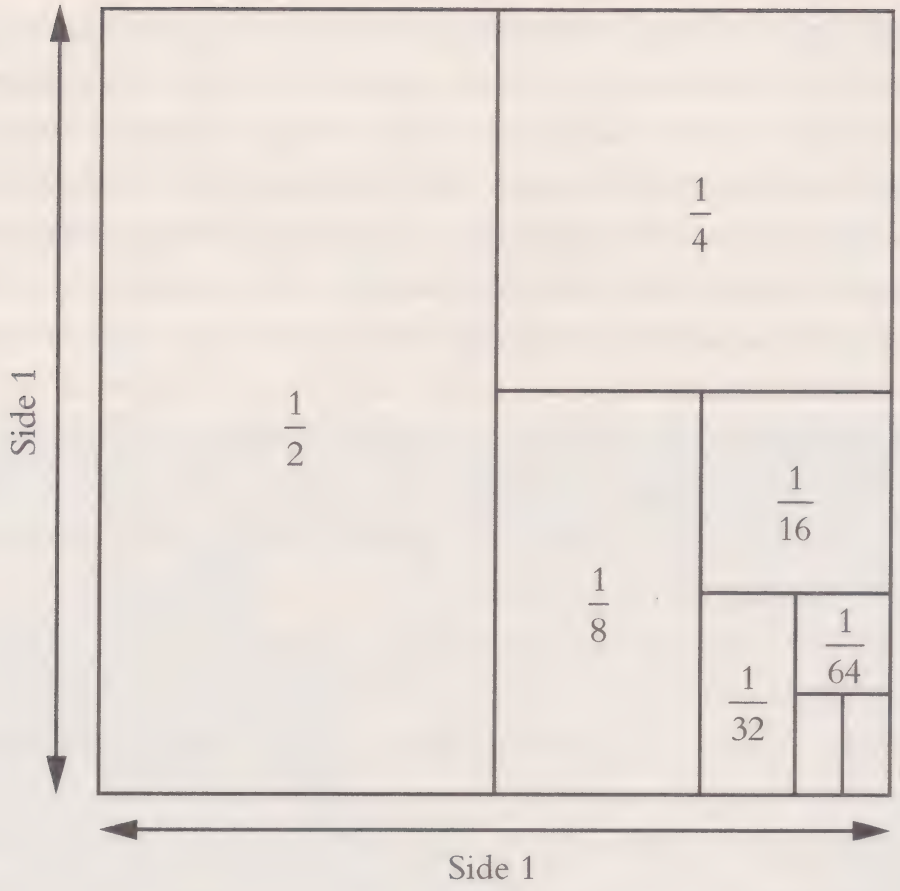
In Miguel de Cervantes' magnum opus, Don Quixote also posed a paradox for his squire.

Quixote had explained that among the sciences that a knight-errant has to know are mathematics, "for at every turn some occasion for them will present itself to him". It was to be so when Sancho Panza was faced with the case of the owner of a stretch of land bordered by a river who forced everyone who wanted to cross the river to first swear on oath where they were going. If they told the truth, they could cross the river, but if they lied they would be immediately hanged. Since the time this law came into being, the judges let nearly everyone cross over the river without problems, until one fine day a man turned up and swore he was going to be hanged. After deliberating about this oath, the judges said: "If we let this man pass free he has sworn falsely and by the law he ought to die; but if we hang him, as he swore he was going to die on that gallows, and therefore swore the truth, but the same law he ought to go free."

It is not much help to us in our studies that, since there were as many reasons for hanging him as there were for not doing so, Sancho Panza recommended the man be set free, as, he said, "It is always better to do good than to do evil." What we do find interesting to note is that the two most successful historical paradoxes, Achilles and the tortoise and the liar paradox, are in fact very different. On the one hand, Zeno's argument for denying Achilles victory over the tortoise is based on an erroneous concept of infinity. Supposing that the advantage is one metre, what the philosopher explains is that Achilles must cover the distance

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\text{etc.}$$

to catch up with the tortoise, as first he must cover half (1/2); then half of the half, that is, a quarter (1/4); then half of a half of a half, in other words, an eighth (1/8), and so on. As there are infinite addends, this distance is infinite, so Achilles would never live long enough to cover it and beat the tortoise. What Zeno didn't know is that the sum of infinite numbers does not have to be infinite, provided that they become progressively smaller with a certain speed. In fact, a beautiful geometric argument produced by Nicholas Oresme (1323-1382) shows that Zeno's addition not only is not infinite, but that the result is 1, exactly the advantage that Achilles gave the tortoise. Zeno's paradox is therefore simply an erroneous concept of infinite sums.



The graph with which Nicholas Oresme demonstrated in the 14th century that the sum involved in the paradox of Achilles and the tortoise is not infinite.

The same thing does not occur in the liar paradox: 'This sentence is false' is a statement about which it cannot be decided if it is true or false, because either answer implies the contrary. As the Greek logician Chrysippus of Soli pointed out, those who composed the liar paradox "completely disregard the meaning of words; they only produce sounds, without expressing anything". The first natural reaction is to attribute the contradiction to the fact that the statement refers to itself, but the self-reference alone is not enough to explain the paradox, because the statements "This sentence is true" or "This sentence is from the book entitled *The Folly of Reason. Mathematical Logic and its Paradoxes*" are also self-references, but even so they present no problem whatsoever.

Another, somewhat more devious, solution was to raise the issue of whether it might be that the concept of truth, like the concept of a set, is easy to use but difficult to define. That was the opinion of Alfred Tarski (1902-1983), who in 1933 published an article of more than two hundred pages, written in Polish, in which

he provided the first formal definition of 'truth'. Despite the length of his text, Tarski was not intending to give the word 'true' a new meaning, but to capture mathematically the Aristotelian notion of truth as a correspondence between what is stated about reality and what reality is. In the same way that 'snow is white' if and only if snow is white, a proposition *P* is true in a theory if and only if on interpreting *P* in the structure referred to by the theory, *P* is true. But, then, in which structure should a sentence like 'This sentence is false' be interpreted? As we shall see in Chapter 4, only Kurt Gödel would manage to answer the question.

At the end of the day, the Russell, the Achilles and the tortoise and the liar paradoxes had solutions, but on the way many more paradoxes appeared. In 1905, a secondary-school teacher in Dijon, Jules Richard, had discovered a paradox in relation to Cantor's diagonal argument. A year later, a young librarian at the Bodleian Library in Oxford – not the one who spent his nights compiling the catalogue of all the catalogues that did not list themselves – had simplified Richard's paradox by imagining what would happen if only 15 words could be used to describe a natural number. As the number of expressions formed by 15 words is finite, in this way we shall only be able to describe a finite quantity of numbers. Among all the numbers that we are not able to describe with this method there will be one that is the smallest; let's call it *n*. But then, *n* is 'the smallest number that we cannot describe in fewer than 15 words', and this description has... 12 words!

How can we be sure that paradoxes will not go on reproducing themselves like a virus? The infinite, self-reference and excessively vague concepts were the source of contradiction, but it was not all the self-referencing that gave rise to them, nor did it seem possible to eliminate the infinite from mathematics, nor did we have a compass that pointed out the imprecise concepts. In the next chapter we shall look at the strategy used by the most brilliant mathematician of his generation, David Hilbert, to try and completely wipe out paradoxes.

Chapter 3

Hilbert's Programme

God exists since mathematics is consistent,
and the Devil exists since we cannot prove it.
Attributed to André Weil

"Who among us would not be happy to lift the veil behind which is hidden the future; to gaze at the coming developments of our science and at the secrets of its development in the centuries to come?"

A new century was beginning, and thousands of people were strolling round the stands at the Universal Exhibition of Paris under the scorching August sun. Meanwhile, David Hilbert was beginning his speech at an amphitheatre in the Sorbonne, speaking for the first time at an International Congress of Mathematicians not about what had been proved but about what still remained to be discovered. No one doubted that Hilbert was the greatest mathematician of his generation, but the lecture had been relegated to one of the congress's secondary sections where there was also a study on ancient Japanese geometry and the proposal to adopt a common scientific language for all countries. Hilbert had, of course, been invited to give one of the plenary speeches at the Paris meeting, but the German mathematician took so long deciding what his theme would be that the organisers finally had to exclude him from the programme. On seeing him approach the stand wearing his distinctive spectacles, the audience wondered what David Hilbert had been up to all this time.

"History teaches the continuity of the development of science. We know that every age has its own problems, which the following age either solves or casts aside as profitless and replaces by new ones." Hilbert was convinced that the only engine generating progress in mathematics was problem solving. That's why, on addressing the audience at the Sorbonne, the leader of the Göttingen school made a great point of what it meant to really solve a problem. That is, of the importance of finding an argument which, starting from a finite number of hypotheses formulated in exact

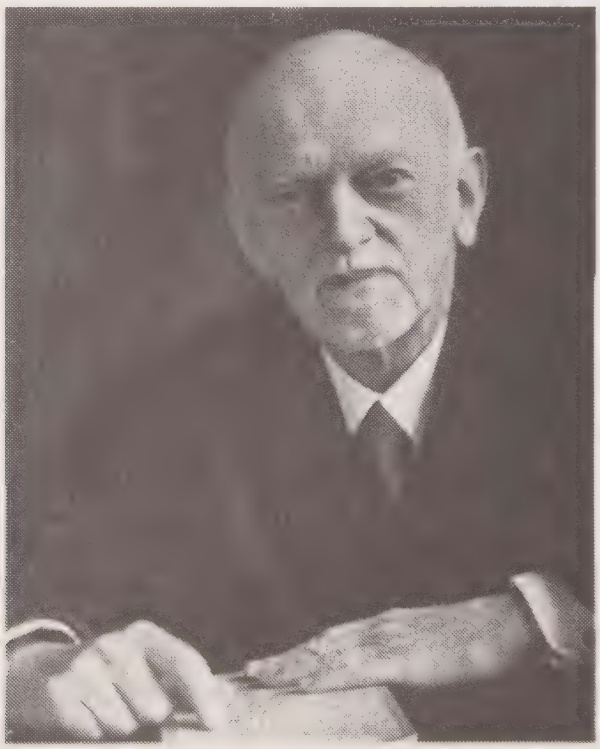
terms, would reach its conclusion after a finite number of rigorous deductions. To illustrate these ideas, Hilbert chose 23 questions, which in his opinion would mark the course of the 20th century's mathematical explorers, though he did not have time to speak about all of them. Thanks to the writings of his friends, fellow mathematicians Hermann Minkowski (1864-1909) and Adolf Hurwitz (1859-1919), we know how much effort Hilbert put into deciding which problems to speak about in Paris. However, there was one problem that he never at any time had any doubts about including. The second problem on his list wondered – apparently in all innocence – are arithmetical axioms non-contradictory?

THE CARDINALITY OF CONTINUUM PROBLEM

In the previous chapter we saw that one of Georg Cantor's great discoveries was to prove that not all infinite sets have the same size. Indeed, the diagonal argument shows that there are fewer natural numbers than infinite successions of 0s and 1s. The first problem on Hilbert's list was to give an affirmative or negative answer to the question of whether there is any set with a cardinality greater than that of the natural numbers but smaller than that of the successions of 0s and 1s. Thanks to the work of Kurt Gödel (1940) and the Stanford University mathematician Paul Cohen (1963), we now know that this question cannot be proved or refuted based on the usual axiomatisation of set theory.

When Hilbert gave his speech, on 8 August 1900, the first paradoxes of set theory had already emerged, but it was to be almost a year before Russell discovered the contradiction that would set all the alarm bells ringing. News of the paradox of the set of all the sets which are not members of themselves would quickly spread and shock the mathematics fraternity. In England, Whitehead would announce the end of 'glad confident mornings'; in Germany, Frege would add a resigned appendix to his *Foundations of Arithmetic*, and in France a victorious Henri Poincaré, the enemy of mathematical logic, would repeat: "Formal logic is no longer sterile: it fathers contradictions". If any mathematician was expected to make a brilliant reply, it was David Hilbert, seen by many as the nearest thing to a reincarnation of Euclid. Such a reputation sprung from 1899 when he gave up the study of number theory to publish an axiomatisation of geometry, which pioneered the modern point of view. However, Hilbert did not bother to find an answer that would go down in history

like those of Whitehead, Frege and Poincaré: It was not necessary when one knew how to eliminate paradoxes from mathematics.



David Hilbert was the man tipped to put an end to paradoxes.

The formalist programme

The solution proposed by Hilbert consisted of two stages. Firstly it was necessary to completely formalise arithmetic, which meant translating all its contents into a formal system. This process had to be carried out with the utmost rigour, but the logicians could not stop there. The first stage would have to be followed by a second, in which it must be proved that the formalisation was consistent. Unlike the case of Caesar's wife, it was not enough for mathematics to appear consistent, but they also had to be consistent, and in order to prove the consistency, Hilbert proposed a set of techniques that he called 'metamathematics'. The reader will be wondering – quite rightly – what difference there is between the axiomatic systems that we have looked at up to now, and the formal systems that Hilbert was seeking for arithmetic. Although the two concepts seem very similar, there is a fundamental characteristic that distinguishes formal systems. Within them, any affirmation has been translated

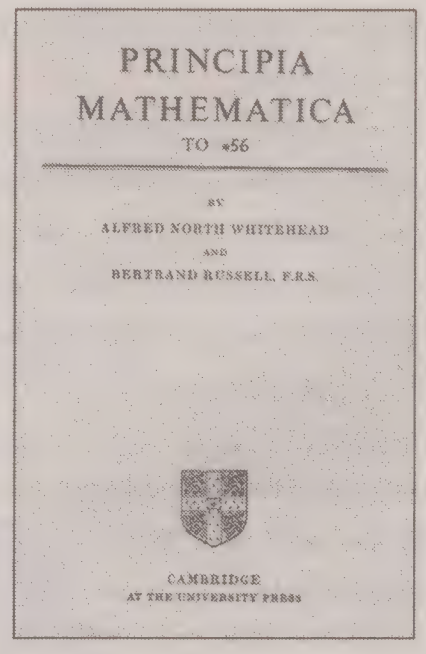
into a series of symbols of an artificial language, which appear devoid of meaning. What Hilbert was aiming at can be understood very clearly in the light of his correspondence, in which he explains, for example, that geometry does not change if, instead of 'point', 'straight line' and 'plane', we write 'love', 'law' and 'chimney sweep'. As a result, for a formalist, 'chapter three' and 'chapter 3' are two different expressions whose only relationship is the fact that they both begin with the same word.

The basis of a formal system as Hilbert saw it is a set of 'primitive symbols' L, which represent the alphabet of our language. Based on them, we can generate 'formulae', which are simply finite chains of symbols constructed in accordance with a series of grammatical rules. If, for example, the language contains opening and closing brackets, one of the rules might be that, for every opening bracket, there must be a closing bracket further to the right. As well as specifying the alphabet, to define a formal system some axioms and rules of deduction are also necessary. 'Axioms' are formulae like the rest, with the only difference that we have given them a privileged role. As we pointed out in the first chapter, the choice of axioms is one of the most difficult tasks when starting up a formal system. If we choose too many, we run the risk of them getting mixed up with the other formulae and we will never be able to distinguish them. However, if we select too few there will be formulae that cannot be proved or refuted in the theory. As for the rules of deduction, these are procedures that allow us to obtain formulae based on the existing ones. Axioms and rules of deduction, also called rules of inference, are combined in formal proofs. These are chains of formulae in which each of them is either an axiom or is obtained from the previous ones by applying rules of deduction. As usual, the last formula in a demonstration is called a 'theorem'.

Therefore, the first task for Hilbert's programme was to specify an alphabet, some axioms and some formal rules of deduction for arithmetic. This is the task to which Bertrand Russell and Alfred North Whitehead devoted the three thick volumes of *Principia Mathematica*, published between 1910 and 1913. In actual fact, Russell and Whitehead's proposal, which immediately came to be known as logicism, went further than the formalist programme. They were not satisfied with just formalising arithmetic but wanted to reduce it to logics. In other words, define all the concepts of natural number theory based on purely logical notions, and also deduce all arithmetical theorems from those principles. One of the great successes of 19th century mathematicians had been to construct any kind of numbers starting from the natural numbers, so if Russell and Whitehead were successful in their purpose,

MATHEMATICS DOESN'T PAY

Russell and Whitehead's magnum opus was published by Cambridge University Press. However, the prestigious publisher was not willing to pay more than £300 in publication costs, which came to half the estimated amount. The Royal Society of London, of which Russell was a member, had promised to make up any shortfall, but in the end they only provided £200, and Russell and Whitehead had to pay the remaining £100 out of their own pockets. "A fine business" – Russell would joke years later, "We earned minus 50 pounds each for ten years' work."


PRINCIPIA
MATHEMATICA
TO *56
BY
ALFRED NORTH WHITEHEAD
AND
BERTRAND RUSSELL, F.R.S.
CAMBRIDGE
AT THE UNIVERSITY PRESS

then mathematics and logic would forever go hand in hand along a pathway that was free of contradictions (or that was, at any rate, their hope).

In a simplified version, the formal system for arithmetic proposed by Russell and Whitehead in *Principia Mathematica* is formed by the primitive symbols 0 (the number zero), s (the successor function), $\neg$ (the negation), $\vee$ (the disjunction 'o'), $\exists$ (existence), $=$ (equal) and opening and closing brackets, and then by adding variables x, y, z of type 0, which therefore represent natural numbers as well as variables A, B, C of type 1, that is, sets of natural numbers, and so on as new levels become necessary. An attentive reader may have noticed the lack of other symbols that ought to form part of the language. For instance, in the same way that we have included the quantifier of existence $\exists$ thanks to which we can formalise statements such as 'There is a natural number with the property P', we should add a symbol that means 'for all', as in, 'For all natural numbers the statement P is true. In fact, that universal quantifier does exist and its use is very widespread in mathematics: 'for all' is written $\forall$. We could have added the symbol $\forall$ to the language, but in reality it is not necessary as 'For all natural number the statement P is true' says the same thing as 'There is no natural number for which the statement P is not true'; so the symbol $\forall$ can be reconstructed from the symbols of negation and existence.

The same thing happens with the conjunction 'and'. We represent it with the symbol $\wedge$ but it is redundant if we already have $\vee$ and $\neg$. To prove it, we will make use of three operations from set theory: the complement, the union and the intersection theories.

Given a set A contained in another set B, the *complement* of A in B is the set formed by the elements that belong to B, but not to A. For example, the complement of the vowels $\{a, e, i, o, u\}$ in the alphabet are the consonants. Let's move on now to union and intersection. Given two sets X and Y, their *intersection* $X \cap Y$ is defined as the set of the elements that belong to X and to Y at the same time. For example, if X were the set of even numbers 0, 2, 4, 6, 8, 10... and Y were the set of the multiples of three 0, 3, 6, 9, 12, 15..., to calculate the intersection we would have to look for the common elements, which are 0, 6, 12, 18..., that is, multiples of six. On the other hand, the *union* $X \cup Y$ is the set to which all the elements of X and all the elements Y belong. Following on with the previous example, the list of numbers of the union of X and Y would read 0, 2, 3, 4, 6, 8, 9...

The great similarity between the symbols representing the intersection of two sets ($\cap$) and the conjunction of two affirmations ($\wedge$) on one hand, and the union of two sets ($\cup$) and the disjunction of two affirmations ($\vee$), on the other, is no accident. If we associate to properties P and Q the sets of numbers that fulfil the properties, let's say X and Y, then the numbers that fulfil P and Q simultaneously are the elements of intersection $X \cap Y$, and the numbers that verify P or Q. At least one of the properties is a member of the union $X \cup Y$. The complement of a set, meanwhile, corresponds to the negation of a statement. In 1880 the British mathematician and philosopher John Venn created some diagrams that are very useful for representing the complement, the union and the intersection of two sets. By making use of them we can demonstrate that the conjunction of the properties P and Q is equivalent to the negation of the disjunction of the negations of P and of Q, or, expressed in a different way: $P \wedge Q = \neg (\neg P \vee \neg Q)$ which allows reconstruction of $\wedge$ from $\vee$ and $\neg$.

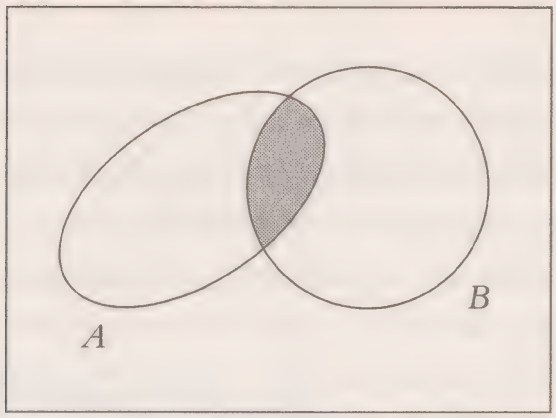


Fig. 1: The intersection of two sets, corresponding to the conjunction $P \wedge Q$.

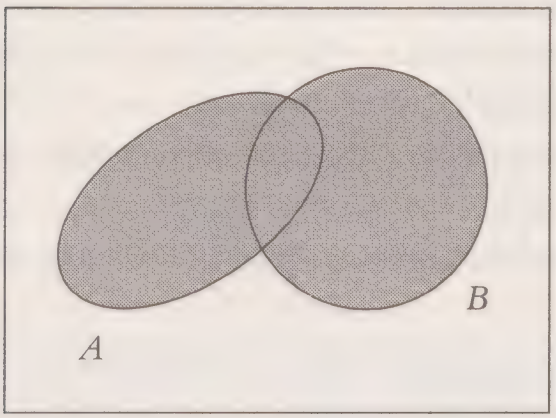


Fig. 2: The union of two sets corresponding to the disjunction $P \vee Q$.

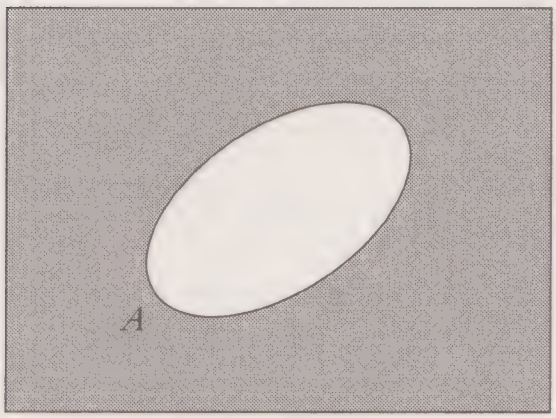


Fig. 3: The complementary of a set corresponding to the negation $\neg P$.

Venn diagrams showing operations of intersection (fig. 1), union (fig. 2) and complement (fig. 3) in set theory.

Having pointed out how to represent 'for all' and the conjunction of two statements, let's see how some of Peano's axioms are translated to the formal system of arithmetic. Remember that the first one stated that "Zero is a natural number". This need not be translated as we have included the symbol 0 in our language. Let's go on to the second: "Each natural number has its successor". First of all, we have to note that two variables are present in this axiom: the natural number in question, which we shall denote by x, and its successor, which we shall call y. By remembering that the successor of a number is written by putting the letter s in front of the number, the relationship between x and y is expressed by means of the formula $y = sx$. In other words, 'y is equal to the successor of x'. The next step is to take note that 'each natural number' is the same as 'for each natural number', and that, in this context, 'has' means 'there is'. This transforms the axiom into 'For all natural number x there is a natural number y such that $y = sx$'. With the symbol $\forall$ at our disposal the work is complete, and the axiom would read $\forall x \exists y (y = sx)$, where we have used brackets to enclose the property fulfilled by the numbers x and y. As that isn't the case, one last translation has to be made: 'For every natural number x, there

PEANO'S FOURTH AXIOM

Let's translate Peano's fourth axiom, which states that "Two different numbers have different successors", into the formal system of arithmetic. As before, the first thing that we have to do is to identify the variables that intervene, which in this case are two natural numbers x and y. What the axiom says is that x and y cannot be different at the same time as their successors coinciding, in other words: There are no numbers x and y such that:

1. x is different from y;
2. the successor of x is equal to the successor of y.

If the symbol of conjunction formed part of the language, the axiom would be written in the following way:

$$\neg \exists x \neg \exists y (\neg (x = y) \wedge (sx = sy)).$$

As that is not how it is, we have to express it in accordance with the negation and the disjunction. Since to negate a statement twice is equivalent to affirming it, Peano's fourth axiom becomes:

$$\neg \exists x \neg \exists y (\neg ((x = y) \vee (\neg (sx = sy)))).$$

is a natural number y such that $y = sx$' says the same thing as 'There is no natural number x such that there is no natural number y such that $y = sx$', Peano's second axiom is written: $\neg \exists x \neg \exists y (y = sx)$. After this detailed explanation, the reader can see that Peano's third axiom, "Zero is not the successor of any natural number", corresponds to the expression $\neg \exists x (sx = 0)$.

From language to metalanguage

Thanks to the process that we have just described, arithmetic has been emptied of significance and reduced to its formal skeleton. Axioms now do not describe anything, but are just chains of abstract symbols, and proofs have become exercises in combinatorics. But it is still possible to state propositions with meaning. For instance, we can say, 'Peano's second axiom is longer than the third', that 'The quantifier of existence appears twice in Peano's third axiom' or that 'The formula $\neg (0 = 1)$ is a theorem of arithmetic'. The important thing is that no longer is it a case of expressions formalised in language L, but of sentences written in English which refer to the formulae of L. Their protagonists are no longer the numbers but the propositions that speak about the numbers, in such a way that, when we state them, we have crossed the boundary of mathematics and leapt right into the dominions of *metamathematics*. The leap is identical to the one produced when one of the characters in a novel suddenly begins to write another novel. In the same way that literature sometimes changes into metaliterature, mathematics can sometimes switch into metamathematics.

One of Hilbert's fundamental contributions was to clearly distinguish the linguistic levels to which the different statements belonged. Let's imagine a Spanish class in which the teacher uses English to explain the shades of meaning of some word. At that moment there are two languages at play: Spanish, the language the students want to learn, and English, which is functioning as a tool. The same thing happens in a sentence such as 'The formula $\neg \exists x \neg \exists y (y = sx)$ is longer than the formula $\neg \exists x (sx = 0)$', which combines chains of symbols of language L with the expressions 'formula' and 'be longer', which do not form part of L, but of a *metalanguage* which we use to refer to the formal system from outside, so to speak. The terms 'zero', 'successor' or 'equal' are allowed in language L, where they are written 0, s e $=$, respectively, but the words 'formula', 'proof' and 'true' belong to a metalanguage that L cannot interpret. Therefore, when arithmetic is formalised, all these statements lose their meaning within arithmetic.

But what has this got to do with paradoxes? Let's not forget that the ultimate aim of Hilbert's programme was to eradicate them from mathematics. As we pointed out in the previous chapter, many paradoxes stem from phenomena of self-reference, which are possible in natural languages but have no reason to be so in the artificial language of formal systems. While it seemed quite reasonable to hear the Russell paradox stated in the English language that there were two classes of sets – those that were members of themselves and those which were not – a formal system would have immediately spotted that the membership relationship applied to two variables of the same type was breaching the grammatical rules. The case of the 'This sentence is false' liar paradox is even more extreme. For it to be taken seriously, not only would self-reference have to be admitted in the formal system, but it would also have to be possible for the property of 'being true' to be expressed in the language, as well as in the metalanguage. Hilbert's hope was that the two situations would never occur at the same time as long as a proper job was done of formalising arithmetic.

However, just hoping wasn't enough, and it is here where the second phase of Hilbert's programme was to come into play, which proposed to put an end to the crisis in the foundations of mathematics by *metamathematically* demonstrating the consistency of formalised arithmetic. Only in this way could future mathematicians be sure that they would never come across contradictions. As if that were not enough, not all methods would be permitted in this proof. Use was to be made of only the soundest, which Hilbert – without ever giving much of an explanation why – christened with the German word *finit*, which would later become *finitary*. These finitary methods had to eliminate any reasoning that was not tangible. They were not to accept, for instance, proofs by *reductio ad absurdum*, an argument that Euclid had used to prove that there are infinite prime numbers or that the square root of two cannot be obtained by dividing two natural numbers. The first step in a proof by *reductio ad absurdum* is to negate the assertion that you want to prove. If, for example, it is a case of proving that there are infinite prime numbers, then the first hypothesis will be that there is only a finite number. Based on this supposition, correct and logical deductions are made until an absurd affirmation is reached, such as, for example, that a theorem that has been independently proved is not verified. All the intermediate reasoning is valid, so the only explanation possible for the absurdity is that the initial hypothesis was false, and thus we have proved just what we intended to. Often, when proving that a certain mathematical object exists, the solution to an equation for example, rather than construct it, it is easier to see that if it didn't exist then we would reach an absurdity. The same thing happens with

metamathematics. We might not be able to prove a statement like 'The formula P is provable' by specifically finding a demonstration of P, but we could by reasoning that if it did not exist it would produce a contradiction. However, for Hilbert such procedures were not sound enough.

David Hilbert was not the only one who rejected such non-constructive methods. Together with logicism and formalism, another response to set theory paradoxes

POINCARÉ VERSUS HILBERT

Henri Poincaré (1854–1912), dubbed by some historians 'the last universalist', hated those who wanted to reduce mathematics to formal relationships between symbols. In 1899, when Hilbert published his *Foundations of Geometry*, Poincaré wrote a long piece in which he criticised the German mathematician for trying 'to get mathematics to work like a pianola'. Some years later, without yet having a very clear idea of the distinction between language and metalanguage, Hilbert would attempt to prove the consistency of arithmetic by using the principle of induction, in other words, Peano's fifth axiom. Ever vigilant, Poincaré pointed out the vicious circle into which Hilbert had fallen on trying to prove the consistency of arithmetic by making use of arithmetic's most important axiom. It was no use Hilbert defending himself by saying that his method was not induction but *metainduction*, because Poincaré was right. The German mathematician would later come to acknowledge this with the help of his pupil Hermann Weyl (1885–1955).

Henri Poincaré.

had been developed, one that involved eliminating all use of the infinite. For the intuitionists, all mathematical objects were the product of the human mind, their existence being equivalent to the possibility of constructing them. The followers of this movement differentiated between the *potential* infinity, which is that of the sets that can be augmented as much as one wants, and the *actual* infinity of completed totalities. They admitted that natural numbers were potentially infinite, as to any finite set of the form $\{0, 1, 2, \ldots, n\}$ still more numbers could be added, but not that this could be referred to all the natural numbers at the same time. Neither did

the intuitionists accept the axiom of the excluded third, according to which, if a statement is not true, then its negation is true. By rejecting that principle the followers of this movement also had to reject all the mathematical theorems that had been proved by it. In fact, even the theory's founder, the Danish mathematician L.E.J. Brouwer (1881-1966) himself, had to reject many of his previous brilliant results because he had used the axiom of the excluded third in them.

Another good example of the techniques that the intuitionists wanted to eliminate from mathematics was the axiom of choice that Ernst Zermelo had proposed for the set theory. Given a finite or infinite collection of sets, this principle allowed an element to be chosen from each one of them, so forming a new set. Those who could not accept the actual infinite were hardly going to like this way of choosing elements as if by magic and without following any explicit rule.

In a series of articles published between 1904 and 1927, David Hilbert gave more and more details of his strategy for replacing all mathematical demonstrations by proofs carried out by finitary methods and for culminating his programme by proving the consistency of arithmetic in the most rigorous and reliable way possible. What the leader of the Göttingen school was not expecting was that a young Austrian, who had begun by studying physics but had soon become more attracted to mathematics, was going to discover while working on developing the formalist programme that Hilbert's dream was impossible, and worse, he was going to do it by using finitary methods!

Chapter 4

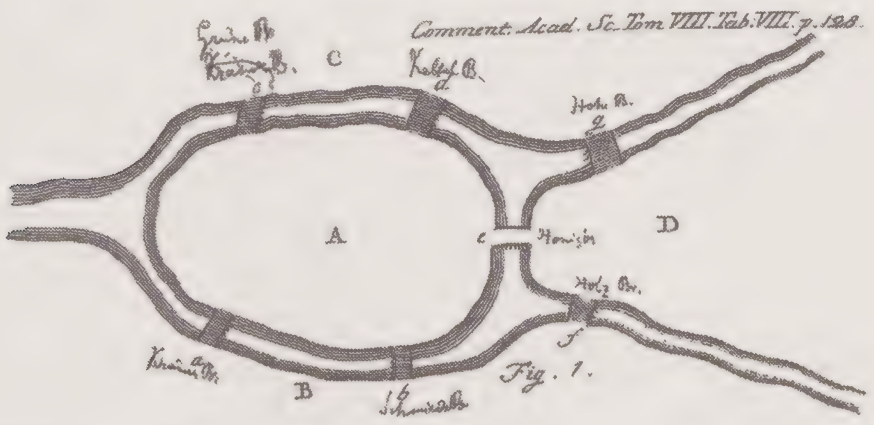
Gödel's Theorems

If controversies were to arise there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates and to say to each other: "Let us calculate".

Gottfried Leibniz

The streets of Königsberg had seen everything. Ever since the city's seven bridges had been built, the inhabitants had wondered if it would be possible to cross them all just once and then return to the starting point. No one had been able to do it, but neither had anyone been able to prove it was impossible until, in 1735, the Swiss mathematician Leonhard Euler invented graph theory and provided a negative answer to the question.

Forty years later, Immanuel Kant would stroll over those same bridges trying to determine the limits of the capacity of pure reason. This fortunate events alone, to which would be added the fact that the city (now Kaliningrad in Russia) was the birthplace of David Hilbert, must have been sufficient justification for a society defending empirical philosophy, which collaborated with the Vienna Circle, to decide to hold the Conference on Epistemology of the Exact Sciences in Königsberg from 5–7 September 1930.



A diagram showing Leonhard Euler's original solution to the Bridges of Königsberg problem.

The meeting's objective was to discover to what degree a solution had been found over the first few decades of the century to the crisis in the foundations of mathematics that had been created by Russell and his paradox. The plenary speakers were chosen for being those who had helped most in recent times to develop the three main solutions to the crisis. They were logicism, which held that all mathematics can be reduced to logic; formalism, whose great success was to distinguish language from metalanguage; and intuitionism, which aimed to expel the infinite from mathematics. The rest of the programme was reserved for participants to present their latest discoveries and for them to enjoy long chats in the city's cafés, which were no doubt not as good as Vienna's, but pleasant all the same.

The Austrian logician Kurt Gödel had been invited to speak on his doctoral thesis, which still left the door open to all-powerful mathematics. However, in the months that had gone by between the start of his brilliant career and the Königsberg congress, Gödel had moved on with his research and become convinced that the dream of a generation of logicians was impossible. There was nothing to indicate this while he gave his speech, but in the final minutes of the roundtable that

The University of Königsberg, commonly known as the Albertina, around the year 1900.

DIALOGUE FROM *THE OXFORD MURDERS* (ÁLEX DE LA IGLESIA/JORGE GUERRICAECHEVARRÍA, 2008)

Sheldon: Oh, I forgot that I was talking to the champion of universal logic. Thank you. You and the police think that it's possible to demonstrate the truth. On the basis of certain axioms and by using valid reasoning you can reach a valid conclusion, isn't that so?

Martin: As sure as today is Wednesday.

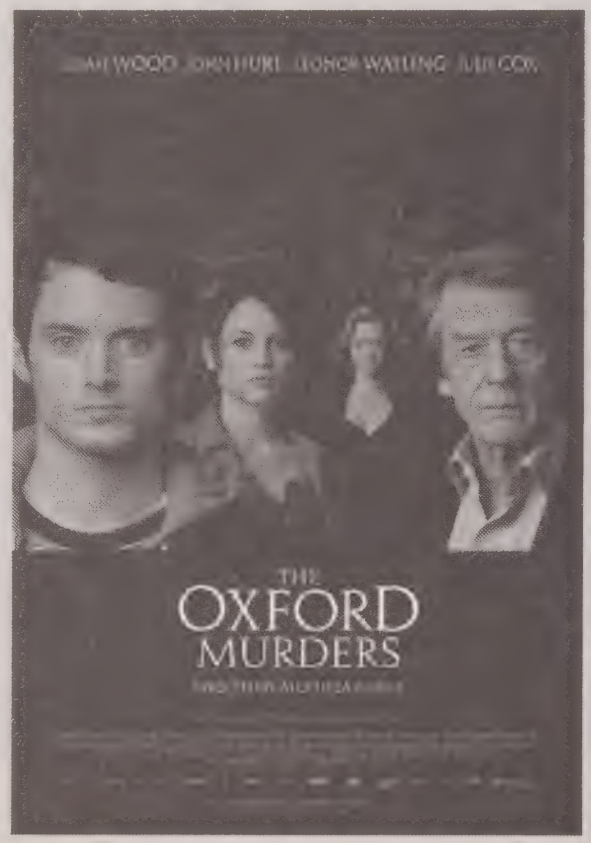
Sheldon: If I were to say "All Britons are liars"? True, false or impossible to prove?

Martin: All right. There are mathematical formulations that can neither be proved nor refuted starting from axioms. Indeterminable propositions.

Sheldon: Exactly. Gödel's incompleteness theorem. So, even in your world of mathematical purity, there are still things that can never be proven.

Martin: Yes, I know, but that is not the case here.

Sheldon: There's a gap, there is a gulf between what is true and what is provable. We can never be sure of all the facts about a phenomenon, and to lack just one could change everything.



concluded the meeting on the following day he at last dared to announce that he had "examples of contentually true propositions that could not be proved from the axioms". Like the end of a novel whose protagonist hangs himself from a nail that appeared back on the first page, Kurt Gödel's words caught the audience so unawares that there was hardly any debate on the issue and they were not even recorded in the proceedings of the meeting.

Not all those present failed to realise that the discreet young man in round spectacles was about to change the course of logic with his rather incomprehensible comment. Among them was John von Neumann who, thanks to his legendary mental speed, immediately realised what Gödel might be referring to and asked him for more details when the conference closed. Among the many universities where he had attended courses, von Neumann had studied with Hilbert in Göttingen, and although he had published several articles following up the master's work, he soon began to doubt that the finitary methods proposed by formalism would be effective in demonstrating the consistency of mathematics. In his youth von Neumann had obtained some results that pointed in that direction and which had

Apart from his contributions to logic, John von Neumann (János Neumann in his native Hungary) also carried out important work in quantum physics.

encouraged him to work incessantly on the issue. One night he dreamt that he had overcome the last obstacle, woke up in alarm and continued thinking about the problem until the next day. When he went to bed the next evening there were still a few loose ends. That night he again dreamt that he had reached the solution, but when he tried to write it out in his waking hours he found an error in the argument, and finally decided to leave it and move on to other matters.

Now, after arriving at Königsberg as the star guest, John von Neumann was suddenly seeing how a second figure was robbing him of fame by announcing what he himself might well have dreamt on that third night. Back home from Königsberg, Hilbert's former helper discovered that, if the Austrian's research was correct, then the consistency of arithmetic could not be proven within arithmetic itself. He informed him of this on 20 November 1930, but unluckily for him three days earlier Gödel had submitted to the journal *Monatshefte für Mathematik und Physik* a manuscript entitled 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I', in which the same result also appeared. Instead of infuriating him, the incident aroused the admiration of von Neumann. When the article was published in the spring of 1931, he interrupted his lecture series in Berlin to explain it, and 20 years later he would still refer to the achievement as "a landmark which will remain visible far in space and time".

David Hilbert was also present in Königsberg, not at the Conference on Epistemology of the Exact Sciences but at a meeting of the society of German scientists who had invited him to give the speech 'Logic and the Understanding of Nature' the day following Gödel's announcement. Although Hilbert and Gödel never had a conversation with each other, the Austrian logician is known to have stayed in Königsberg for several days following the meeting, so it's not improbable that he would have been among the audience who heard Hilbert proclaim more passionately than ever that unsolvable problems do not exist in mathematics: "We must not believe in those who, today, with philosophical bearing and deliberative tone, prophesy the fall of culture and accept the *ignorabimus*[1]. Because for us there is no *ignorabimus,* and in my opinion none whatever in natural sciences. In opposition

1 Abbreviation of the Latin *ignoramus et ignorabimus*, in other words, 'we don't know and we'll never know', which the German physicist Emil du Bois-Reymond coined in 1872 to express his pessimism regarding the limits of scientific knowledge.

to the foolish *ignorabimus*, our slogan shall be: We must know – we will know!" His powerful voice was still echoing when Hilbert found out that the events of Königsberg were putting his programme in danger.

Incompleteness theorems

Prior to Gödel's announcement at the conference, the progress of Hilbert's programme had been giving cause for hope. The first requirement, formalising mathematics, seemed to have been successfully completed in Russell and Whitehead's *Principia Mathematica,* and several logicians were working on proving the consistency of classical formal systems, beginning with arithmetic. Although in the introduction to his doctoral thesis Gödel had already suggested the possibility that there existed "true sentences that cannot be deduced in the system in question", his aim was not to put an end to Hilbert's dream but to prove the validity of the programme. However, the intellectual spirit of the era was pointing in another direction. As a result of Gauss' geometry research, it had been concluded that it was impossible to draw a perfect map of the Earth. Évariste Galois (1811-1832), for his part, had demonstrated that hardly any algebraic equation can be solved with simple methods, while Werner Heisenberg (1901-1976) had just established a new limit for science with his uncertainty principle, according to which it is impossible to accurately measure the position and speed of electrons at the same instant.

Through his theorems, Gödel was to make everyone aware of the intrinsic limitations of the axiomatic method. If in the first chapter we explained that the attributes that make a formal system so formidable are consistency (i.e. that it does not create contradictions), recursiveness (that the axioms can be recognised amongst the rest of the statements) and completeness (that truth coincides with provability), the Austrian logician would show how, in the case of arithmetic, these three conditions are incompatible. According to Gödel, no recursive and consistent axiomatisation of arithmetic can be complete. Put another way, there will always be only some true properties of the numbers that we cannot prove with axioms. This is the contents of Gödel's *incompleteness theorem*, usually referred to by experts as Gödel's first theorem, for he still found the force to prove a second theorem, asserting that the statement 'Arithmetic is consistent' is an example of these undecidable propositions. That was the same result that von Neumann had managed to deduce after the meeting in Königsberg.

To demonstrate the incompleteness theorem, Gödel changed the liar paradox into an undecidable sentence that showed no contradiction whatsoever. Indeed, part of the theorem's indubitable appeal is this way of living dangerously, just one step away from the paradoxes, but without falling into them. The reader will remember that in the second chapter we saw that one of the formulations of the Epimenides paradox was 'This sentence is false'. If the statement is taken to be true, then it itself says that it is false, while if it is considered as false, it must be true. But now then, what would happen if we looked for the provable instead of for what is true? Let's give the name G (G for Gödel) to the proposition 'This sentence is not provable' and let's suppose that our axiomatic system is consistent. If G were false, as what G says is 'I'm not provable', then G would be provable, but in a consistent system no false statement can be provable, as a contradiction would immediately occur. If G is not false, then it is true, so we have a true sentence that says: 'I am not provable'. Therefore, the moment we suppose that the system is consistent, we can find a true but unprovable sentence; in other words, 'consistent' implies 'incomplete'.

The moment we suppose that the system is consistent... But, what system? Any kind reader who asked herself or himself that question at the end of the previous paragraph, may have thought that it was their own fault for having got lost among so much self-reference and for not knowing to which system we were referring. But they have just asked the crucial question, to which there was no answer before Gödel. Our reasoning shows that the affirmation 'I am not provable' must be true, but it is not a mathematical statement, as we would like it to be, but a metamathematical one, because it does not refer to the subject matter of any theory but to the theories themselves. Gödel's genius consisted of translating some metalanguage expressions to the language of arithmetic, thanks to a system of codification based on prime numbers. After this *gödelisation* of metamathematics, natural numbers led a double life. On one hand they were themselves, what they had always been, but on the other hand they played the role of some formula, which enabled a statement like 'I am not provable', which a priori only had any sense in metalanguage, to become a numeric relationship.

Until we come to a more detailed explanation of Gödel's code, it is enough for us just to be aware that by using this code, a statement equivalent to "I am not provable" could be expressed in arithmetic. If arithmetic had a recursive and consistent set of axioms *S,* then there was a formula G_S which was true but unprovable (here

"ANYTHING THAT'S NOT ON YOUR LIST"

Randall Munroe (b.1984) used to work for NASA until in 2005 he discovered his great talent for making people laugh with science humour. He then began to draw the *xkcd* series, "a webcomic of romance, sarcasm, math and language". It consists of very simply drawn cartoons, and often includes references to results from physics, mathematics and IT. Kurt Gödel has made a number of star appearances, but none so brilliant as the cartoon entitled *Fetishes*, reproduced below.

Randall Munroe during a talk at the Massachusetts Institute of Technology (source: Petehume).

we have used the subindex S to indicate that the constructed sentence depends on the axioms, so that if we changed them we would get a different one). To the ever-present logicians, Gödel offered a choice between two diverging pathways: one of completeness and another of consistency and recursiveness. Even worse, arithmetic was not only incomplete but also incompletable. At the beginning of this book, when we gave the example of the police chief who had just joined the local police force, readers might have protested that his colleagues would have known

if he were married or not if they had just chatted to him a little more. There are incomplete systems that stop being so just with the addition of a handful of axioms. But that is not the case in arithmetic. As well as showing the undecidable sentence G_S, Gödel proved that it is no use incorporating it as an axiom as by applying the method to $T = S + G_S$, which is again a recursive and consistent set of axioms, another proposition is obtained which is true but unprovable G_T. Cutting off one of the Hydra's many heads will never save us from incompleteness.

We promised that we would explain how it is possible to translate the undecidable proposition 'I am not provable' into arithmetic, but before doing so we shall move forward a few steps to the second incompleteness theorem. In the first chapter we said that in inconsistent axiomatic systems any proposition is a theorem. Therefore, the existence of at least one formula that is not a theorem is an unmistakable criterion for knowing when a theory is consistent. If we are able to find an unprovable proposition, we will automatically free ourselves of contradictions. Just one is enough! So, why choose a very complicated one when we have the simplest one of all to hand: $0 = 1$? At the start of the book, we showed how the theorem 'Zero is different from one' was deduced from Peano's axioms. The reader will not find it difficult to see that, even if we choose other axioms, any sensible theory that refers to numbers will distinguish zero from one. In short, to say that arithmetic is consistent is the same as saying that the formula $0 = 1$ is not provable.

We again find ourselves faced with a metalanguage statement, but which by means of *gödelisation* can be transformed into a formula of numbers, which we shall call Con_S (*Con* from consistency and *S* for the system of axioms). In this translation, what the first incompleteness theorem says is that Con_S implies G_S, because if the arithmetic is consistent (that is, if Con_S is true), then G_S is also true. At this point it should be remembered how one of the most powerful rules of deduction, *modus ponens,* works. This rule enables deduction from proofs of logical implication 'If *A* then *B*' and from the statement *A*, of a proof of *B*.

Let's suppose for a moment that the consistency of arithmetic could be proven within arithmetic. Then Con_S would be provable and by putting it together with the proof of the first theorem of incompleteness, $Con_S \rightarrow G_S$, we would deduce by *modus ponens* a proof of G_S. But that is absurd, as G_S is unprovable! The only possible conclusion is that to prove the consistency of arithmetic it is necessary to go outside arithmetic, and that is what the second incompleteness theory says. This is what Gödel himself considered to be a 'surprising corollary' to his research.

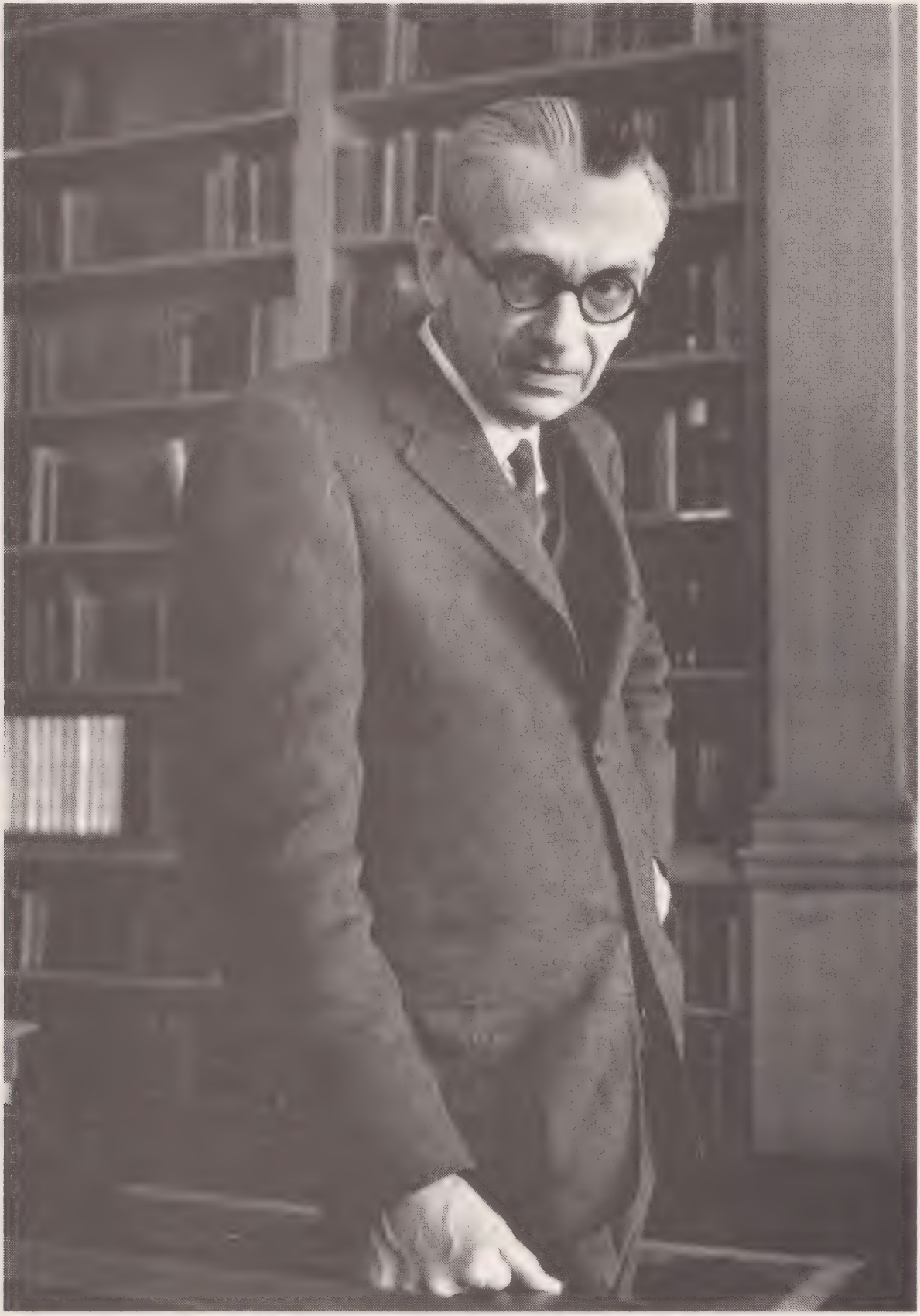
According to Hilbert's programme, to prove the consistency of mathematics it was necessary to start from arithmetic. However, Gödel's second theorem pointed out that, if a proof of the consistency of arithmetic existed, use would necessarily have to be made of techniques that were much more complicated than the finitary methods defended by the formalists. The reader will probably have noticed that Gödel's title *On Formally Undecidable Propositions of* Principia Mathematica *and Related Systems I* indicated there would be second part; the reason is that the article only contains an outline of the second incompleteness theorem. Although everything in it is correct, Gödel never finished writing it all down, which fits in with the image of him as the 'explorer who leaves all the details to others' given by his biographers. It was in fact David Hilbert and his colleague Paul Bernays (1888-1977), to whom Gödel had explained all the subtleties of the proof during a voyage across the Atlantic,

CITIZEN GÖDEL

After fleeing Nazi Germany in 1940, Kurt Gödel took up permanent residence at Princeton University. One of the best known anecdotes on him concerns an event that took place seven years later when he applied to become a US citizen. As happens with all applicants, Gödel had to demonstrate knowledge of US legislation in an examination on the Constitution. In practice, the test was no more than a bureaucratic procedure, but Gödel wanted to thoroughly prepare himself for it and, while doing so, believed he had found some contradictions in logic:

"Up to now, you held German nationality."
"Excuse me, sir, Austrian," replied Gödel.
"Ah, yes, that damn dictator. Fortunately, that's not possible in America."
"On the contrary," broke in Gödel. "I know how!"

But rather than let him carry on, the judge, whom Albert Einstein had already warned that Gödel was no ordinary candidate, took control of the situation and led the procedure through to more routine questions with: "It's not necessary to go too deeply into it." Round about that same time, some logicians had begun to lay the foundations of a theory, deontic logic, the aim of which is precisely to prevent contradictions arising when new laws are being drawn up.



Kurt Gödel in a photograph taken at the Princeton Institute of Advanced Study, New Jersey.

who published the first complete proof on the second incompleteness theorem in 1939. It is a symptom of the healthy moral values in science at that time that it was Hilbert himself who should finish off the details of a theorem that demolished his own over 25 years' work.

Even so, the reception that the incompleteness theorems were given was not quite as good as they deserved. Some mathematicians thought that the undecidable proposition 'I am not provable' was merely a curiosity that would never affect their work, and there were those who, on failing to understand the subtlety that separated the true from the provable, accused Gödel of reproducing the liar paradox. Among them was the sexagenarian Ernst Zermelo, in spite of the fact that he knew more than anyone how hard it is to fight for an idea, as his axiom of choice had brought him immense criticism.

In general terms, the mathematical community was not ready at that time to understand a work that incorporated very innovative techniques in what had always been a minority field. How right was Thomas Kuhn when he pointed out in his study *The Structure of Scientific Revolutions* that in science "novelty emerges only with difficulty, manifested by resistance, against a background provided by expectation". Luckily, when Gödel's article was the translated into English and given more exposure from the 1960s onwards, incompleteness theorems to begin to be acknowledged as the most important advance in logic since the times of Aristotle.

Gödelisation

On 21 June 1851, in one of London's oldest restaurants, Adolf Anderssen, at the time the best chess player in the world, met up with Lionel Kieseritzky, who taught chess at a club in Paris, to play what in the years to come would become known as 'the immortal game'. Impressed by the strategy of Anderssen, who had sacrificed his bishop, queen and two rooks to checkmate him, Kieseritzky immediately wanted to send a description of the game to his club. But instead of beginning 'White: the fifth pawn to the left moves two squares forwards. Black: the pawn in the same column is placed opposite it. White: the third pawn on the right advances two squares. Black: the pawn that moved in the first move takes the last-named piece…', the first symbols of the message were in the style of 'e4 e5 / f4 exf4…'.

All the information on the game took up scarcely three lines, and that was lucky for Kieseritzky! Because if he had used the first method, paying for the telegram would have been an expensive business in the Café de la Régence where he played chess for five francs an hour.

Chess players had found an extremely concise method of condensing all the information on their moves. They had made use, firstly, of an old method of translation, Descartes' analytic geometry, thanks to which each square on the board could be identified by two coordinates: one letter from *a* to *h* to represent the columns, and number from 1-8 to indicate the rows. Except for the pawns, which were left unidentified, each piece was represented by its initial: B for bishop, Q for queen, K for king and R for rook. Only the knight breaks the rules by having a phonetic N. Then other symbols were added, such as x for take, + for check, and ++ for checkmate. In this algebraic protocol, the sequence 'e4 e5 / f4 exf4' said the same thing as 'White moves a pawn to square e4, and black responds by moving another pawn to square e5. Next, white moves a pawn to square f4, which black takes with the pawn that was on square e5.'

Our reason for giving this example is to emphasise how useful codification systems are in many fields both within and without mathematics, where they can transform very complex expressions into easy-to-handle symbols. In the previous chapter we saw how the properties of natural numbers, written in day-to-day language, could be translated into the symbolism of the *Principia Mathematica*. For example, the axiom 'Zero is not the successor of any number' turned into the formula $\neg \exists x\,(sx = 0)$ by means of this system. Gödel, however, needed to go further. To demonstrate the incompleteness theorem it was not enough for him to reduce arithmetic to formulae, but rather he had to be able to condense any formula – even any proof! – into a single number. That was when the logician remembered that in the seminars on the history of philosophy that he had attended when studying at the University of Vienna, Professor Theodor Gomperz had reviewed Louis Couturat's editing of Leibniz's unedited manuscripts, which were published in 1903.

Like his most brilliant predecessors, Leibniz made great efforts to try to put an end to the *confusio linguarum* with which God had punished human vanity in attempting to build a tower that reached the sky. To do so he had imagined a universal language that reduced all human thought, irrespective of the language in which it was formulated, to a list of primitive ideas, each of which had a prime

number assigned to it. By using this list, and in the same way that composite numbers are formed, the characters for derived ideas could be calculated, and it would always be possible to "extract the primitive notions of which they are composed". If, for example, the numbers 3 and 5 correspond to the concepts of water and stillness, then the compound idea of a lake could be expressed by the product $3 \cdot 5$. Reciprocally, if we were told that the concept of lake admits the *character* 15, we would decompose 15 into prime factors and, by looking up the primitive ideas associated with the numbers 3 and 5 in the list, we would conclude that a lake is simply still water. So, to find out if an affirmation of the type 'A is B' is true, it would suffice to see that character B divides character A, and 'there would no more need of disputation between two philosophers than between two accountants'. Leibniz's ambitious programme, discovered two centuries after his death, never came to be implemented, but it gave Gödel ideas on how to translate metalanguage into arithmetic.

Remember that prime numbers are those that are only divisible by 1 and by themselves. For example, 5 is a prime because neither 2 nor 3 nor 4 can divide it; but 6 isn't prime, because 2 and 3 can divide into it. The first prime numbers are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31... and by a *reductio ad absurdum* argument, the ones so hated by the intuitionists, it can be proved that the list continues indefinitely. Most of the efforts in physics in the second half of the 20th century have focused on identifying the elementary particles of matter, those that cannot be divided into simpler ones. However, mathematicians have known since the days of Euclid that the elementary particles of arithmetic are the prime numbers. Indeed, on picking any natural number n there are two possibilities. Either n is prime, and so we have finished, or there is some other number other than 1 and n that divides it. If, for example, the value of n is 23, we would be in the first situation, but if n is equal to 30, then it is divisible by 2.

Let's suppose, therefore, that the starting number is not prime; then we can decompose it as a product: $n = a \cdot b$ (in our case, $30 = 2 \cdot 15$). We have now obtained two numbers to which we can apply the process again: if they are both primes, then we have finished, but if one of them isn't, we write it again as the product of the factors. Continuing with the same example, 2 is prime, so there is nothing to do there, but 15 can still be decomposed as $15 = 3 \cdot 5$, so $30 = 2 \cdot 3 \cdot 5$. As 2, 3 and 5 are prime numbers, the game is over. In general, either we find a prime factor, or the terms that appear get smaller and smaller, which ensures that the process will

come to an end sooner or later. In this way we have proved the *fundamental theorem of arithmetic*, which says that any number can be decomposed as a product of prime factors which may be repeated. For example: $77{,}220 = 2 \cdot 2 \cdot 3 \cdot 3 \cdot 3 \cdot 5 \cdot 11 \cdot 13$, and in that case the abbreviation $77{,}220 = 2^2 \cdot 3^3 \cdot 5 \cdot 11 \cdot 13$ is used, where the exponents indicate the number of times that each prime appears.

The fundamental theorem of arithmetic says something more vigorous: Not only does a decomposition of this type exist for any natural number but, what's more, it is unique except for the order of the factors. That is to say, we might be able to write 77,220 in another way, for example, as $77{,}220 = 5 \cdot 2^2 \cdot 11 \cdot 3^3 \cdot 13$, but in the new decomposition the same numbers will appear raised to the same exponents.

In the previous chapter we saw that the arithmetic alphabet is composed of eight symbols: 0 (number zero), s (the successor function), $\neg$ (the negation), $\vee$ (the conjunction 'or'), $\exists$ (existence), $=$ (equals) plus the opening and closing brackets, and we also have variables x, y, z which stand for the numbers we are going to study. As the first stage of codification, Gödel proposed making each symbol correspond to a number from 1 to 8, and the first three primes over 8 to the variables x, y, z, as can be seen in this table:

0	s	$\neg$	$\vee$	$\exists$	$=$	(	)	x	y	z
1	2	3	4	5	6	7	8	11	13	17

Once a number has been given to arithmetic's 'primitive ideas', codifying a formula is very simple. First you count the number of symbols that appear in it (with repetitions) and you choose the same number of prime numbers; the size of the formula does not matter, because there are an infinity of primes. Next, each prime number is raised to the exponent corresponding to the symbol, in accordance with the previous lexicon, and they are all multiplied among themselves. Let's see it at work with an example, which is worth a thousand explanations.

Peano's third axiom states "Zero is not the successor of any number," which we express as $\neg \exists x\,(sx = 0)$. If we follow the *gödelisation* instructions literally, the first thing that has to be done to transform it into a number is to count the symbols that appear in the formula; there are nine: $\neg$, $\exists$, x, (, s, x, $=$, 0). We therefore

choose the first nine prime numbers, that is: 2, 3, 5, 7, 11, 13, 17, 19 and 23. In accordance with the dictionary, the negation ¬ is associated with number 3, so we have to raise prime 2 to the power of 3, that is: 2^3. In the same way, the quantifier of existence ∃ is shown by number 5, so that prime number 3 has to be raised to power 5, that is: 3^5. By repeating the process, we get 5^{11}, 7^7, 11^2, 13^{11}, 17^6, 19^1 and 23^8, and by multiplying them all it becomes:

$$2^3 \cdot 3^5 \cdot 5^{11} \cdot 7^7 \cdot 11^2 \cdot 13^{11} \cdot 17^6 \cdot 19^1 \cdot 23^8.$$

The method we have just described allows every formula to be codified into a number, which we shall call its Gödel number, but there is nothing stopping us from doing the same thing with proofs. Let's remember that a proof is nothing more than a finite sequence formed by, say, *n* formulae, so it is possible first to codify each of the formulae, then to choose *n* prime numbers, raise them to the Gödel number of each of them and then to take the product. In this way, each arithmetical proof is reduced to a number.

The crucial point is that *gödelisation* is a reversible process. Those who are familiar with chemistry will know that one of the areas of most interest in this science is knowing which reactions can be run in reverse and turned back into their starting points. When a fuel is burned, for example, it is transformed into water vapour and carbon dioxide, the famous greenhouse gas. It is, however, impossible to recover the original fuel from those gases; if it were, the planet's energy problems would have been solved by now! Other chemical reactions, on the other hand, are reversible. For example, passing water vapour over a hot sheet of iron produces hydrogen gas and iron oxide. The iron oxide and hydrogen can then react to re-form pure iron and water vapour.

Thanks to the fundamental theorem of arithmetic, in the *gödelisation* laboratory all reactions are reversible. Let's see why this is true. We will start with the very large number,

304,496,379,203,017,490,604,020,678,113,081,132,612,291,772,080,917,708,
404,389,616,093,394,253,015,558,500,327,468,465,234,375,000.

which we have taken the trouble to write so that the reader can get an idea of big the smallest Gödel numbers are... The fundamental theorem of arithmetic

assures us that it is possible to decompose this figure into its prime factors. If you don't feel like working it out, which is quite a reasonable attitude to take, given the size of the number, you can go to http://www.wolframalpha.com and write the number (without the commas, which are only there to make the number easier to read) preceded by the word 'factor', in the main box. For still bigger numbers the computer could take a long time, but what is important here is that the fundamental theorem of arithmetic guarantees that such a factorisation always exists and that, what's more, it is unique. Luckily, the Internet considers the number 304,496...375,000 to be a small one, and sends back its decomposition in less than a second:

$$2^3 \cdot 3^5 \cdot 5^{11} \cdot 7^3 \cdot 11^5 \cdot 13^{13} \cdot 17^7 \cdot 19^{13} \cdot 23^6 \cdot 29^2 \cdot 31^{11} \cdot 37^8.$$

The only thing left to do is to take note of the exponents and get the symbols from the dictionary. And so we get the formula $\neg \exists x \neg \exists y (y = sx)$, which says that there is no number x such that there is no number y with the property that y is the successor of x. By reformulating the proposition slightly, readers can be assured that we can write it as "Each natural number has a successor", which is Guiseppe Peano's second axiom.

Of course, not all natural numbers are the Gödel number of some formulae, but even if we were given one that did not correspond to any arithmetical expression, we would know how to spot it immediately. For example, $15 = 3 \cdot 5$ is not the Gödel number of any formula, as *gödelisation* makes it necessary for the first primes to appear, without any jumps, and 2 does not appear in the decomposition of 15. The number $1{,}536 = 2^9 \cdot 3$ does not correspond to any arithmetical expression either, because even though in this case the primes appear in order, none of the symbols of the alphabet corresponds to the exponent 9.

To recap, the codification system we have described here enables a number codifying its structure to be linked to every formula (as well as to every proof) of arithmetic. Additionally, this mathematical reaction is reversible in the sense that, by factorising any natural number N, we can decide:

1. If N is the Gödel number or not.
2. If it is, what formula it represents.

GÖDEL IN LITERATURE

In William Boyd's novel *The New Confessions*, the protagonist has just made the *magnum opus* of silent movies, but the film's launch is largely ignored as it coincides with the first films being made with soundtrack. Only Kurt Gödel, in a fleeting appearance, is capable of recognising the film director's great talent.

In another novel published ten years later, *In Search of Klingsor*, written by the Mexican author Jorge Volpi, the girlfriend of the protagonist, a physicist named Francis Bacon, bursts into a seminar that Gödel is giving at the Institute of Advanced Studies and begins to scream at him that he is cheating on her. When the action switches to the back rows, "Professor Gödel announces that he cannot continue with the lecture and begins to sob uncontrollably". His great conflict – the author has von Neumann say – is not the formally undecidable propositions, "but his wild and turbulent love for a prostitute: his own wife". Whilst the portrayal given in *The New Confessions* is true, the scene described by Volpi is as cruel as it is ludicrous.

The writer William Boyd included Gödel in his novel The New Confessions.

Proof of the incompleteness theorems

As we have seen, the inspiration for Gödel's splendid method of codification came from reading Leibniz and, although we have spent a good while on it here, we should remember that it is just a tool to help us reach our goal – to prove that in any recursive and consistent axiomatisation of arithmetic there are propositions that are true but unprovable.

At the beginning of the chapter we pointed out the strategy behind the proof. We had to replace the concept of truth with that of provability in the liar paradox so as to get the statement that says 'I am not provable'. If we do not admit contradictions, the sentence must be true, so it is unprovable. As pointed out then, the greatest difficulty was to find the arithmetical equivalent for this metalinguistic proposition, which does not deal with numbers but with mathematical theories. Well then, at our disposition now we have all the methods for translating it. Below we shall try to explain the most important steps in Gödel's proof in the simplest way possible.

The game consists of translating the sentence 'I am not provable' into arithmetic, so the first question we should deal with is what does it mean when a proposition is said to be provable in the axiomatic system of arithmetic? It means that there is a proof that ends with our statement. In other words, a finite chain of formulae in which each of them is either an axiom or has been deduced from the previous ones by means of the permitted rules of inference.

To find out if a certain sequence of formulae, which we shall call Z, proves the statement X, we must verify that Z is constructed in accordance with the previous rule, and that its last formula is precisely X. By means of the *gödelisation* process, the key idea is to associate X and Z with their Gödel numbers, which we shall label with the small case letters x and z. What we would most like to do is to have a mechanism D (D for demonstration) that would take the natural numbers x and z, and would eventually answer whether the sequence of formulae corresponding to number z is a demonstration of the formula with Gödel number x or not. Therefore, the sentence $D(x, z)$ would be true if Z proved the formula X, and false if it did not.

To take a very elementary example, let's remember that the Gödel number of Peano's second axiom is $2^3 \cdot 3^5 \cdot 5^{11} \cdot 7^3 \cdot 11^5 \cdot 13^{13} \cdot 17^7 \cdot 19^{13} \cdot 23^6 \cdot 29^2 \cdot 31^{11} \cdot 37^8$. As axioms are characterised by being their own proof, if in $D(x, z)$ we replace the values of x and z with that number, the result is true, as the sequence of formulae

with Gödel number z, in this case composed entirely of Peano's second axiom, is a proof of the formula *gödelised* by x; once again Peano's second axiom! If, however, we were to introduce a value of z as number $2^3 \cdot 3^5 \cdot 5^{11} \cdot 7^7 \cdot 11^2 \cdot 13^{11} \cdot 17^6 \cdot 19^1 \cdot 23^8$, the mechanism $D(x, z)$ would say 'false', as the corresponding formula is not a proof of Peano's second formula. The fact that the formula with Gödel number x is provable means that there exists a number z such that the sequence of formulae corresponding to z is a demonstration of the formula associated with x or, in other words, a z such that the statement $D(x, z)$ is true. Consequently, the formula $\exists z\, D(x, z)$, which we shall abbreviate to Dem(x) (Dem from demonstrable), states that the formula with Gödel number x is provable.

To sum up, if D existed, thanks to *gödelisation* all the subtleties of the provability could be reduced to a simple relationship between the natural numbers x and z. And what is the theory that deals with these relationships? Arithmetic!

As the reader will have guessed, the most laborious part of Gödel's article was to prove that a mechanism with these properties did indeed exist. To do so, the Austrian logician needed 46 stages, of which we shall only give an outline. Let's suppose that we take the natural number z, which is the code for some sequence of formulae. Thanks to the fundamental theorem of arithmetic, we can decompose z into its prime factors:

$$z = p_1^{k1} \cdot p_2^{k2} \cdot p_3^{k3} \cdot \ldots \cdot p_n^{kn}.$$

It is not our intention to baffle readers with this notation, which is necessarily somewhat complex. The only thing we have done up to now has been to decompose number z into a series of prime factors which are raised to some exponents. As z codifies a sequence of formulae, each of the exponents will be the Gödel number of one of them. This process enables us to identify the *gödelisation* of each of the formulae on the list, which we have called k_1, k_2, k_3... up to k_n.

Let's just repeat one more time the theme tune that has been running throughout this book: a demonstration is a sequence of formulae in which each of them is either an axiom or is obtained from the previous ones by means of the permitted rules of inference. Therefore, what has to be verified is the following:

First step: the sequence of formulae with Gödel numbers k_1, $k_2 \ldots k_n$ has the structure of a proof. That is to say, the statement corresponding to

each of these numbers is either an axiom or it is deduced from the previous ones by one of the permitted rules of inference.

Second step: the last formula in the sequence is the one that we want to prove.

Let's start at this last stage, which is the simplest. We have been given a formula with Gödel number x, and we want to know if the chain of statements ends in that formula, which is the most basic requirement for it to be a proof. Now then, the previous calculations have enabled us to identify the Gödel numbers of each of the formulae on the list, and the one that corresponds to the last formula is none other than k_n, so it is enough just to see if the numbers x and k_n are equal! No one would deny that finding out if two numbers are equal is a simple task.

Let's go on now with the first stage of this obstacle course by examining formulae with Gödel numbers k_1, k_2... up to k_n, to see if they behave as they should. It is here that it becomes essential for the system of arithmetical axioms to be recursive, which is something that up to now might have appeared rather a whim. Let's remember that a set of axioms S is recursive when it can be verified, in a finite number of steps, whether a proposition is an axiom or not. Therefore, at our disposition we have a formula $A(x)$ (A for axiom) which reads the number x and decides if the corresponding proposition is an axiom or not. So far so good; it will be enough for us to calculate $A(k_1)$, $A(k_2)$... up to $A(k_n)$, which shows which of the statements of the potential proof are axioms. The first formula, corresponding to Gödel number k_1, necessarily has to be an axiom, as there is nothing before it from which it could be deduced. Therefore, if $A(k_1)$ should happen to be false, we would have finished: z is not the Gödel number of a proof. Let's suppose, however, that everything is going well for the time being.

Among the next formulae, shown by numbers $k_2, k_3 \ldots k_n$, some will be axioms and some will not. For those which aren't, it will be necessary to check that they are deduced from the previous ones through permitted rules of inference. In his very detailed work, Gödel shows that for every rule of deduction there is a formula I (I for inference) which takes the first s numbers k_1, k_2... up to k_s and answers 'true' if the formula with Gödel number k_s is deduced from the formulae of Gödel numbers k_1, k_2... up to k_{s-1} (the immediately previous one) by applying the corresponding rule of deduction. For example, $I(k_1, k_2, k_3, k_4)$ will be true if the fourth formula in the chain is deduced from the three previous ones by applying the rule of inference that has been codified by means of the

formula I. In this way, we can apply this process to the formulae that are not axioms and if, for each of them, at least one of the answers to the different rules of deduction is 'true', then the first stage has been concluded successfully, and z is the Gödel number of a proof.

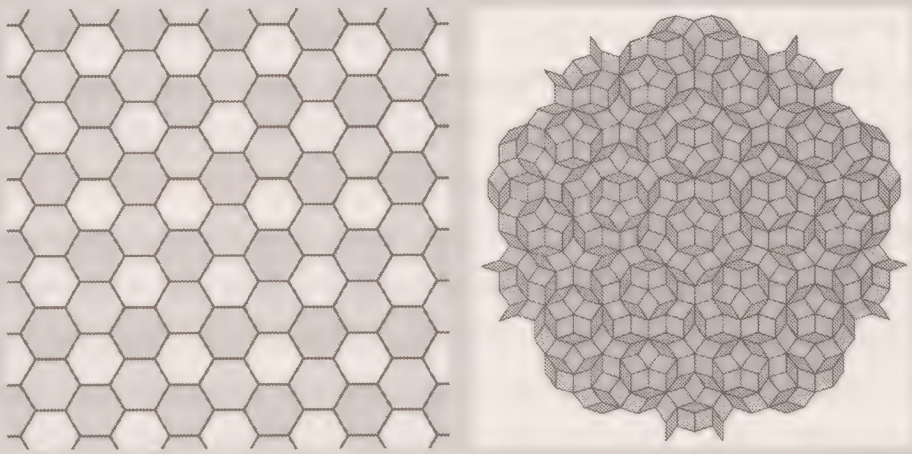
As it is easy to get lost among the technical details, we'll point out the most important factor: what needs to be remembered is that we have proved that there is a process $D(x, z)$ which decides if the sequence of formulae represented by z is a proof of the statement of Gödel number x. To do so, we just need to translate the rules that a proof must follow into relations between numbers, which we have been repeating as if it were the chorus of a song.

So far so good. We have now constructed using arithmetic the statement $\text{Dem}(x)$, which says 'The formula with Gödel number x is provable'. By negating it we get $\neg\,\text{Dem}(x)$, which simply says 'The formula with Gödel number x *is not* provable'. Up to here there is no mystery, but we are getting nearer and nearer to the crucial step. But before we reach those acrobatics it is necessary to remember that the statement 'arithmetic is consistent', which appears in the second incompleteness theorem, is equivalent to stating that 'The formula $0 = 1$ is not provable'. Bearing in mind that one is the successor of zero, that is, $1 = \text{s}\,0$, we invite the reader to verify that the Gödel number of the formula $0 = 1$ is 255,150. Therefore, the proposition $\neg\,\text{Dem}\,(255{,}150)$, translated into the language of arithmetic, states that 'The formula with Gödel number 255,150 is not provable', that is, 'The formula $0 = 1$ is not provable', which is the same as 'Arithmetic is consistent'. The statement $\neg\,\text{Dem}(x)$ kills two birds with one stone.

The important thing about the expression $\neg\,\text{Dem}\,(x)$ is that it is no longer a statement in day-to-day language but, instead, it is now a formula in arithmetic, in which only the symbols $0, \text{s}, \neg, \vee, \exists, =, (,)$ and some variables appear. The letters 'Dem' are just a way to abbreviate it, because to write it out would be extremely complicated and take up many pages; but if we wanted to do so, we could write it by exclusively using the characters of arithmetic. That's why we have worked so hard! I am sure that readers now know what to do whenever they come across a formula written that way – gödelise it! So, with $\neg\,\text{Dem}\,(x)$, let's associate its Gödel number, which we will call d. This number may be so gigantic that there is not enough ink in the whole world to write it, but our philosophy has always been that size does not matter; what matters is that it is a number.

THE INCOMPLETENESS OF TESSELLATIONS

A tessellation on a plane is a way of covering it with some type of tile in such a manner that the tiles neither leave any gaps nor overlap. Islamic art offers extremely beautiful examples of tessellations, but we can also find them in nature: for example, bees tessellate their honeycombs in hexagons, the optimum method of doing it. However, not all tessellations need to be so regular. There could be others that are *aperiodic*, in which it is not possible to find any symmetry. In the 1960s, the logician Hao Wang (1921-1995) discovered that if a question on tiling on the plane were undecidable in the same sense that the sentence 'I am not provable' could neither be proved nor refuted, then those non-periodic ways of tiling the plane would exist. As he thought this possibility completely absurd, Wang concluded that his problem had to be decidable. Some years later, however, one of his students proved that with 20,426 different tiles it was possible to tessellate the plane in a non-periodic fashion. Little by little, that number has been reduced to only two different tiles.



On the left, a regular tessellation formed by a single type of regular polygon, similar to a honeycomb; on the right, an example of non-periodic tessellation.

All the structure of the proposition 'The formula with Gödel number x is not provable' is contained in just a single number: d. The parameter x is not fixed; instead, it can take any value. And if it can take any value, why not be awkward and choose x equals d? We would then get the statement $\neg \mathrm{Dem}(d)$, which states that 'The formula with Gödel number d is not provable'. However, as d is in turn the Gödel number of the proposition 'The formula with Gödel

number x is not provable', $\neg$ Dem ($\acute{d}$) turns into 'The formula *The formula with Gödel number* x *is not provable* is not provable'. If we put the proposition in the mouth of the formula with Gödel number x, what is being stated is nothing more and nothing less than 'I am not provable'[2].

What the theorem does not say

The end of the argument that we have just given proves that no consistent and recursive set of axioms in arithmetic can be complete, reproduces the scene in which, after taking a class in logic, many pupils return home sobbing: "Mummy, I'll never be a logician!" while the rest, *the happy few* wear a satisfied grin from ear to ear. We would without doubt prefer for the reader to be among the latter group. Though we may not have been successful in that endeavour, even those who at this stage feel like screaming "Mummy, I'll never be a logician!" or have angrily thrown the book out with the rubbish, will understand that the theorems we have dealt with here have nothing at all to do with a sentence of the type: "Ever since Gödel showed that there does not exist a proof of the consistency of Peano's arithmetic that is formalisable within the theory, political scientists had the means for understanding why it was necessary to mummify Lenin and display him to comrades in a mausoleum".

It is undoubtedly true that the author of that quote, the French essayist Régis Debray, (b. 1940) is noted for his vivid imagination and not for being ignorant. He studied philosophy with Louis Althusser at the École Nórmale Supérieure de Paris. At one time held prisoner in Bolivia, he was freed after an international campaign which united such motley characters as Jean-Paul Sartre and Pope Paul VI. During the moments he was not occupied with politics, he began to write what now comes to about 50 books, among them *The Scribe: The Genesis of the Politician*, from which we took the extract on Lenin's mummy and the comrades.

The case of Régis Debray is not unique. Other intellectuals such as philosophers Gilles Deleuze and Julia Kristeva, psychoanalyst Jacques Lacan

[2] The nature of this book prevents us from being completely rigorous and including formulae of recognition of free variables, substitution and generalisation used by Gödel in his article. We believe, however, that all the essential components of the proof have been included here.

and the architect Paul Virilio have been swept along in recent times by what the French philosopher Jacques Bouveresse called the "prodigies and dizziness of the analogy", and from a very technical result in logics deduced general conclusions that have no relationship at all with mathematics but whose pseudo-scientific wrappings no doubt impress their readers.

Horatio used to say that the voice, once let out, cannot return. Apart from the opinions that can be seen in this book, the reader also has the possibility to consult the original work of Kristeva, Debray, Lacan, Deleuze and Virilio to make up his or her own mind if their ideas are just more examples of the dangers of dealing with subjects about which the author knows nothing or, on the contrary, they are the best reflection of the enormous power of seduction of some theorems which, as John von Neumann said, will always remain visible far in space and time. From here onwards, we shall focus only on those who certainly did know what they were talking about, and it is on that point where one of the most brilliant men in history enters the arena: Alan Mathison Turing.

Chapter 5

Turing Machines

What can I hope?
Immanuel Kant

"Eur..." Betty waited impatiently for the rattling rotors to stop so she could read the rest of the message. "Europe..." More than five years had gone by since the day she found out that the magazine that brightened up her hours as a servant for one of London's richest families was holding a crossword competition. "Europe wi..." Every day she would remember her surprise when she got the news of the prize, and her hesitation before asking for a week off work. "Europe will never..." Then in her mind she would relive the journey she had made together with the other puzzle enthusiasts and picture the vivid memory of the silhouette of Bletchley Park mansion appearing outlined against the grey sky that autumn day. "Europe will never be..." She feared forgetting some detail of a story she would tell everyone as soon as the war ended. R-u-s-s-i-a-n. The last word had taken a while to come out, but Betty could now celebrate a new triumph for the Allies: "Europe will never be Russian." It was 15 April 1945, and that was Adolf Hitler addressing the high command of the Nazi Party.

They were not the only ones who were aware of the dictator's delirium two weeks before he committed suicide: without Himmler even suspecting it, ten thousand people simultaneously read his correspondence with Hitler in a little village eighty miles from London, a place with good communications by train and road, but remote enough to avoid bombings. It was there that in 1939 the Government Code and Cypher School had been set up, its mission consisting of deciphering the instructions that the Nazis encoded with their Enigma machine, the most perfect ciphering machine built at the time. Engineer Arthur Scherbius had conceived of the idea in 1918 to make commercial transactions more secure, but in view of its military potential the German navy soon acquired the rights to it and spent the next ten years perfecting it. When the Wehrmacht's troops invaded Poland at the beginning of September 1939, Enigma's cryptographic methods had become so

sophisticated that the possibility that anyone would be able to decipher its messages was not even considered.

Only the joint effort of a team made up of mathematicians, physicists and translators, backed by a group of women who had been selected through the secret crossword test, could take on this diabolical device which, by means of a system of electrical impulses sent to a series of rotors, transformed the same letter written twice into different symbols. Disguised as pirates, as if they were a bunch of bored aristocrats in search of a bit of wartime fun, in 1939 the first codebreakers took over the barrack huts that had been built beside the Victorian mansion. It was essential that no one in the village nearby should be aware of the crucial task that was to be carried out at Station X, as the facility came to be known. This was where the Allies sent all the messages gathered at the front. Winston Churchill himself referred to Bletchley Park as "my goose that lays golden eggs but never cackles".

On the right, Nazi soldiers codifying their messages by means of an Enigma machine, an example of which can be seen on the left.

Above, the offices at Bletchley Park where the Enigma code was translated. Below, the mansion as it is now.

The Poles had discovered a singularity about Enigma that made it less secure than the Nazis thought: every letter, whatever its position, was always codified by means of a different one. However, many problems had to be solved between the discovery of this first clue and the moment five days before the D-Day landings at Normandy, when the whole Bletchley Park team celebrated deciphering a message in which Hitler gave assurances that troops would land in Calais, nearly three hundred kilometres north-east of Arromanches, the true location. The landing might not even have come about without the information on the position of the Nazi submarines that they managed to decipher at Station X, a surprising achievement bearing in mind that in 1939 the team did not even have an Enigma machine on which to try out their hypotheses.

By working day and night in 8-hour shifts, the Bletchley Park codebreakers managed to build a prototype that was identical to the one the Nazis used, but the

enterprise would never have been successful had it not been for the intervention of a young English mathematician whom many Cambridge University students compared to a Greek god – a *deus ex machina* recruited to win the war. Without Alan Turing (1912-1954) it would not have been easy to see that, sooner or later, all the messages spoke about the weather, and it was at that sentence where the decoding of the message should begin. For the first time, the phatic function of language, that is, the type used when starting communication, was shown as being essential.

Another of the mathematician's proposals was to build a great computer, the Bombe, which would enable them to simulate what was happening in ten Enigma machines simultaneously. If Turing was able to see further than his colleagues it was not because of his years of study in the latest state-of-the-art laboratory, but because for a long time he had explored the boundaries of what he considered to be the most beautiful human creation: Gödel's theorem.

A slate statue of Alan Turing, the work of the British sculptor Stephen Kettle, together with a portrait of the mathematician kept in the National Museum of Computing at Bletchley Park (source: Jon Callas).

DIALOGUE FROM *BREAKING THE CODE* (HERBERT WISE/HUGH WHITEMORE, 1996)

Dilly Knox: I've been furnished with some details of your work Mr. Turing, most of which – I have to tell you – I find almost totally incomprehensible.

Turing: That's hardly surprising.

Dilly Knox: I used to be rather good at mathematics when I was younger, but this is well... baffling. Now, for instance this thing here: 'On computable numbers with an application to the Entscheidungsproblem'. Can you tell me something about it?

Turing: Tell you what?

Dilly Knox: Well... anything, a few words of explanation, in general terms.

Turing: A few words of explanation?

Dilly Knox: Yes.

Turing: In general terms?

Dilly Knox: If possible...

Turing: Well... it's about right or wrong... in general terms. It's a technical paper in mathematical logic but it's also about the difficulty of telling right from wrong. You see, people think that – well, most people – think that in mathematics we always know what is right and what is wrong. Not so!, not anymore! It's a problem that has occupied mathematicians for forty or fifty years – I mean – how do you tell? Right from wrong, you know? [...]

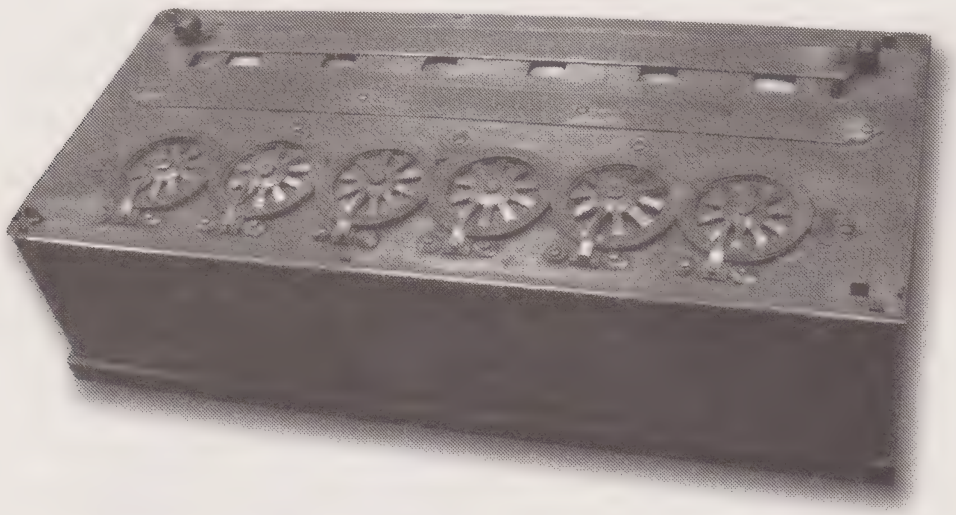
Dilly Knox: Well, I see – well I don't – but I see something, I think, the originality of your thinking is clearly remarkable, and I'm sure you'll make an invaluable member of our team, group, call it what you will.

Thinking like a machine

Far from being isolated events, the building of the Bombe and Colossus – the first programmable computer and also a product of Bletchley Park – followed a line of continuity going back at least to the second decade of the 17th century, when the German astronomer Wilhelm Schickard (1592-1635) made the first 'calculating clock', a mechanical device that could add, subtract, multiply and divide. He would be followed by Blaise Pascal (1623-1662) with his calculator, which he had begun to design at the age of 19 to help in the work of his father, who had been appointed

tax collector for the city of Rouen. Marketed under the name of Pascaline, the new machine was a smash hit in the salons of the aristocracy, where scientists and nobles were fascinated by it. It was there that it was studied for the first time by Gottfried Leibniz (1646-1716). Convinced as he was that "it is unworthy of excellent men to lose hours like slaves in the labour of calculation", it is hardly surprising that he was excited by the Pascaline and immediately wanted to improve on it. His dream was to make a machine capable of recognising all true statements.

At the beginning of the 19th century, Pascal's and Leibniz's calculators inspired the English mathematician Charles Babbage (1791-1871) and his pupil Ada Byron



The Pascaline was the world's first calculator, designed by the Frenchman Blaise Pascal.

(1815-1852) to make the first theoretical reflections on computation. With the aim of building an Analytical Engine, Babbage and Byron set out the basic elements needed to operate in any IT process. Firstly, there has to be a program indicating the operations that have to be carried out. It consists of a series of instructions which, starting from a set of data that we call 'input', enables a result to be calculated which is returned to the user as 'output'. For instance, the input of the program 'multiply' is pairs of numbers such as (2, 3) and the output is their products, in this case, $2 \cdot 3 = 6$. For the program – which we shall often call 'algorithm' – to actually function we will also need a processor that obeys the instructions, plus a memory to store the input data, the instructions and all the intervening calculations. In the case of Babbage's analytical engine, the input was introduced by means of cards that had been perforated in a Jacquard loom, which was initially designed to automatically weave patterns marked out by holes punched in cards.

Ada Byron's father was the great English poet Lord Byron, and her mother was Annabella Milbanke, whom Lord Byron called 'the Princess of Parallelograms' as she had studied algebra and geometry with a Cambridge professor. After giving birth to Ada, Annabella left her husband and, determined that the little girl would not follow in the poet's footsteps, introduced her to the study of science as soon as she could. At the age of 17 Ada met Charles Babbage at a dinner given by her friend Mary Somerville, who had been her tutor and who always encouraged her to continue studying mathematics. Shortly afterwards, Ada would explain to Babbage how to calculate Bernoulli numbers by using a system of perforated cards, a problem that was much more ambitious, from a mathematical point of view, than those the inventor of the analytical engine had solved up to then. With her method of 'weaving pure algebra', Byron not only wrote the first IT program in history, but also showed that to solve a problem of an algorithm type it was not always necessary to start from scratch. A handful of basic operations was repeated in nearly all the problems, so it would often suffice to combine the existing cards. That is what computer experts today call subroutines.

What distinguished Ada Byron from Charles Babbage is the same thing that would enable Alan Turing to lay the rigorous foundations of the Theory of Computation with his article 'On computable numbers, with an application to the Entscheidungsproblem', published in 1937 in the *Proceedings of the London Mathematical Society*. While Babbage was still convinced on his death bed that if he lived just a few more years the fruits of the analytical engine would be seen all

On the left, a commemorative stamp issued on the centenary of the birth of Charles Babbage. Above, gardens dedicated to Ada Byron in Barcelona (photo: Ana Navarro Durán).

over the world, both Byron and Turing had realised that it was necessary to make much more progress in the theoretical field before the first computer could ever be built. One of the questions that most needed to be studied was precisely which problems could be solved by the machine and which could

BERNOULLI NUMBERS

One of the best known anecdotes on Carl Friedrich Gauss tells of when one day his schoolteacher at primary school wanted to take it easy while he had his pupils do a long maths exercise by adding up the numbers from 1 to 100. What Mr Büttner wasn't expecting was that little Gauss would find the answer almost immediately by using a method that would just as easily work to add up the numbers from 1 to 1,000. Let's set, then, the number n that we want to reach. The idea Gauss had was to write the addition $1+2+\ldots+n$ the other way round and make use of the symmetry of the terms as we show here:

$$1+2+\ldots+(n-1)+n$$
$$n+(n-1)+\ldots+2+1.$$

Readers will not find it difficult to see that if we pair each term with the one that is below, the result is always $n+1$. As the process is repeated n times, we shall get $n(n+1)$ as a result. But, look out! By doing it this way, we have added each number twice, once in the first row and again in the second one. It is necessary, therefore, to divide by two:

$$1+2+\ldots+n=\frac{n(n+1)}{2}.$$

One might wonder whether, on replacing the addition of the n first numbers by the n first squares, it is possible to get similar formulae. Using a method that is a little more sophisticated than the previous one, it can be proven that

$$1^2+2^2+\ldots+n^2=\frac{1}{3}n^3+\frac{1}{2}n^2+\frac{1}{6}n,$$

and that the sum of the n first cubes is calculated by means of the formula

$$1^3+2^3+\ldots+n^3=\frac{1}{4}n^4+\frac{1}{2}n^3+\frac{1}{4}n^2.$$

In general, the k-th Bernoulli number is related to the coefficients that appear on writing the addition of the n first powers of order k as a polynomial in the variable n. They are numbers that are easy to define with words but difficult to calculate explicitly. That's why Ada Byron's algorithm was such a great step forward.

not. Something similar is happening today with quantum computation, the theoretical achievements of which are far ahead of the effective design of the first quantum computer.

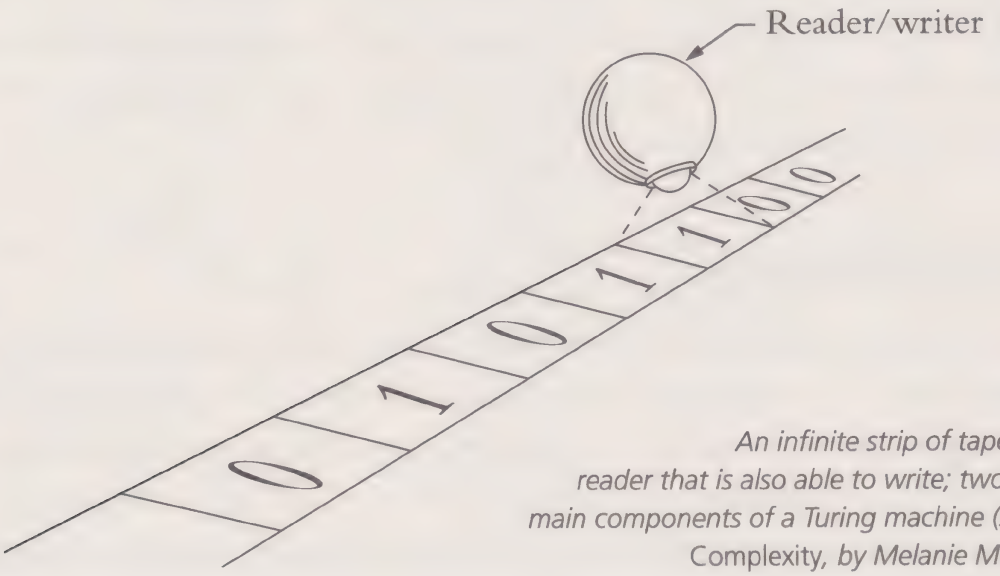
Turing's brilliant idea for exploring the limits of action of future computers was to consider the question of what thinking like a machine really means. It is easy to see that a computer could not have the intelligence and imagination of human beings, which allow humans to cope with surprises and unknown situations. On the other hand, machines do not get tired or bored when they have to carry out long and tiring calculations. They never have a bad day. To distinguish the problems that a computer would not be able to solve due to its technical limitations (for example, because the program we designed requires the age of the Universe to run) from those that are unsolvable on account of the actual conditions of the statement, Turing imagined an ideal computer, with infinite memory and running time. What this machine of Turing's could not do would also be too much for the most powerful computers of the future, and the English mathematician's method would set the limits for what we can expect from computers.

Computable functions

The first success of Turing's research was to define the concept of computable function. From here onwards, every time we say function it shall be understood as a function defined over the natural numbers, which takes natural values. Let's remember that a function is nothing more than a way of associating with each number another number, which we shall call its image. The reader can think of functions – but only if this provides encouragement to carry on reading the chapter – as a machine that shapes the raw material which is put into it. So, our function will convert number 3 into another number which we shall call $f(3)$, the f being function. The process of obtaining $f(n)$ from n may consist of a series of algebraic operations or of a more complex verbal description. For example, if the function were that which associates to each number its successor, which, as we saw at the beginning of the book, appears in Peano's axioms, then we could write $f(n) = n + 1$, and the result would be $f(3) = 3 + 1 = 4$. If, on the other hand, the function were to determine the prime number that occupies the nth position, then $f(3)$ would be 5, and $f(4)$ would be 7, because the prime numbers are 2, 3, 5, 7, 11... In this case we have a description in words, but not a simple formula to express the value of the function at each point.

The image of the machine could give a false impression, perhaps making readers believe that ideal machine of Turing's which we mentioned would be able to compute any function imaginable. On the contrary, the hidden operations between the input of number n and the output of $f(n)$ might be so complicated that not even Turing's machine could carry them out. If we want to distinguish between both situations, it is now time to explain details of the workings of the machines that Alan Turing imagined when he was little more than 20 years old.

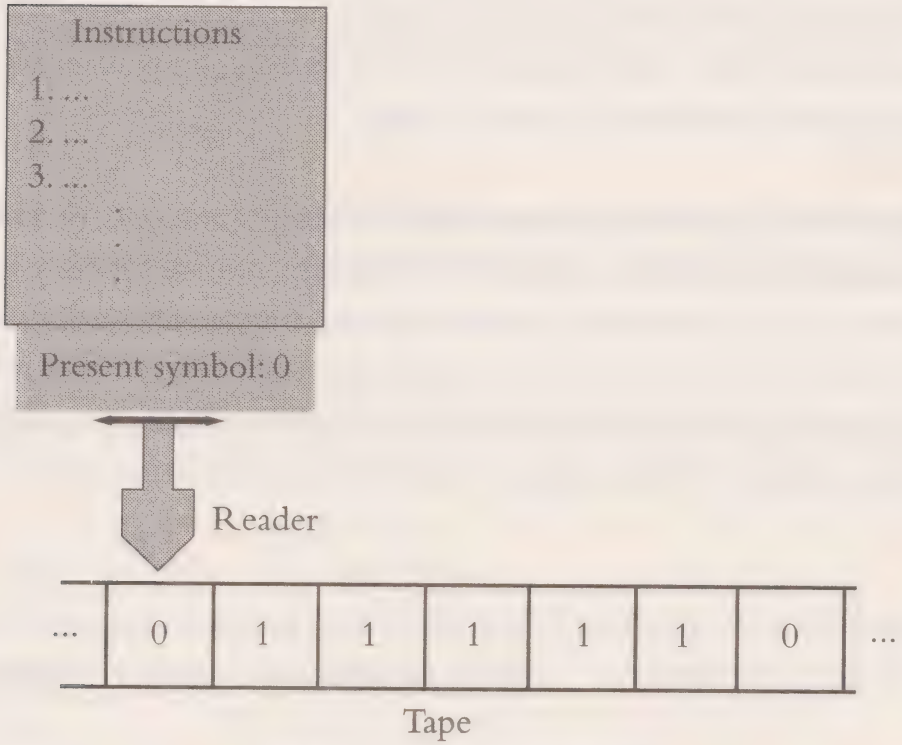
The first element is an infinite strip of tape to the right and to the left – remember that we are talking about an ideal machine – which is divided into cells which each contain only one symbol, let's say either 0 or 1. They correspond, as we know, to the two possible values of truth. The Turing machine's second element is a reader able to detect if the number written in a certain cell is a 0 or a 1 and to write on it.



An infinite strip of tape and a reader that is also able to write; two of the main components of a Turing machine (source: Complexity, by Melanie Mitchell).

After reading it, the device can respond in five different ways: erase the number that was there and write a 0; replace what is written there with a 1; move to the right; move to the left (to make it possible for these two operations to be carried out it is essential for the tape to be infinite); or simply to stop, that is, not to respond to the reading of the symbol at all. The sequence of these actions is controlled by a finite chain of instructions that indicate to the machine how it must respond in each possible situation. For example, the first instruction might be: 'If the symbol seen is 1, move to the left and go to the third instruction.' All the instructions follow this pattern:

Instruction number __: if the reader meets the symbol __, then run operation __ and go to instruction number __.



The workings of a Turing machine
(source: Complexity, *by Melanie Mitchell).*

As we said, the instructions are numbered, starting with 1. The symbols that appear are 0 and 1, and the possible operations are to write a 0 (0), write a 1 (1), go right (R), go left (L) or to halt (H). This enables the instructions to be expressed in symbols through the four data that intervene in each one. So, if the first instruction were 'If the symbol seen is 1, move to the left and go to the third instruction,' it would be enough to write (#1, 1, L, #3). By now, readers will have realised that for each number, two instructions are needed: one to indicate what to do in the event that the symbol seen is a 0, and the other to explain how to react when it meets a 1. In the preceding example, if the third instruction only indicated what to do if a 0 is seen, but if the following symbol were a 1, the machine would not know how to continue. One possible solution is to establish that when there is not a precise instruction, the machine – which has no imagination to enable it to continue by itself – should halt. However, so that the explanation is clearer we shall explicitly explain what action is taken in all possible cases. Let's look at a very simple example, that of the Turing T machine formed by the three commands given below:

Instruction #1: if the symbol seen is 0, write 1 and go to instruction #3.
Instruction #1: if the symbol seen is 1, move to the right and go to instruction #2.
Instruction #2: if the symbol seen is 0, write a 1 and go to instruction #3.
Instruction #2: if the symbol seen is 1, stop.
Instruction #3: if the symbol seen is 0, write a 1 and go to instruction #1.
Instruction #3: if the symbol seen is 1, stop.

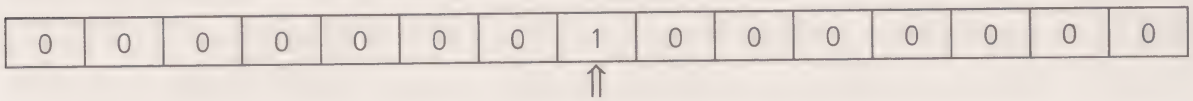
When this Turing machine is being codified through the system described in the previous paragraph, the problem arises of what should be done when the machine halts, as in this case the instruction does not give any further order. The simplest solution is to add a 0 at the end: in this way there is no possible doubt because, even if the Turing machine tries to find the instruction 0, no order bears this number. By using this trick, the following sequence of instructions contains all the information of T:

(#1, 0, 1, #3) (#1, 1, R, #2) (#2, 0, 1, #3)

(#2, 1, H, #0) (#3, 0, 1, #1) (#3, 1, H, #0)

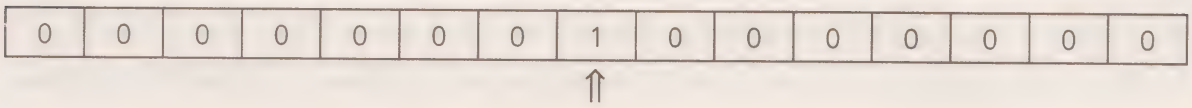
Now let's see how the program reacts if we enter a tape full of 0s as input. The arrow indicates where the Turing machine reader is situated at each moment.

The program will begin to run the first instruction. As the symbol that the reader finds is a 0 and the order is "If the symbol seen is 0, write 1 and go to instruction #3", it is enough to replace the 0 with a 1 and to see what the third order says:

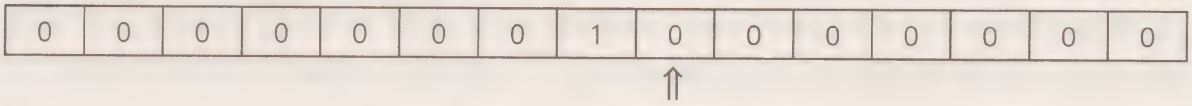


Now, instruction #3 has two parts: the first one indicates that if we read a 0, we have to write a 1 and come back to instruction #1, but according to the second one, if the machine sees the symbol 1 then it must halt. As this is the case, the program has finished running. Therefore, when a tape full of 0s is entered, T halts after writing a 1 at the starting point.

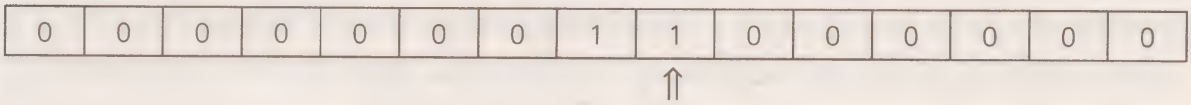
Turing machines are easy, aren't they? Let's see what happens if this time we apply the program to the tape that we have just obtained. The input is, therefore:



We shall begin by applying the first instruction. Because what we read is a 1, we have to move to the right and go to the second instruction. There's no mystery there!



At this point, instruction #2 determines that, on seeing a 0, the Turing machine has to replace it with a 1, and then go on to the third instruction. Let's obey the instructions:



Once more, the instruction #3 tells T to halt when it sees the symbol 1, so the program has finished running, and the output is a tape with two 1s between an infinity of 0s, with the reader situated over the symbol further to the right. If we start up the Turing machine again, the new result will be a tape with three 1s instead of two, so what T calculates is simply the function $f(n) = n + 1$. In general terms, a function is computable if there is a Turing machine that calculates each of its values.

Let's suppose that the natural number n is codified, as it is in the previous example, by the input of a tape comprising n 1s among an infinity of 0s to the right and to the left, with the reader situated on the last of them. A function f will be computable if there exists a Turing machine that, by entering any value of n in this way, outputs the metamorphosis $f(n)$. What we have demonstrated is that the function 'add one' is computable in the Turing sense. As, in order to calculate the function $f(n) = n + 2$, it is enough to apply the same set of instructions twice,

and to compute $f(n) = n + 3$, it is enough to repeat the process three times, and so on, addition is a computable operation. Multiplication is too, since just as we were taught at school, multiplying 3 and 5 means to add number 5 three times, or five times number 3: "The order of the factors does not change the product...".

We were saying that a function is computable if a Turing machine exists that calculates each one of its values, but that does not mean we always know how to find it. For example, let's take a function that only admits as *input* and as *output* the values 0 and 1. It is therefore enough to specify $f(0)$, which may be equal to 0 or to 1, and $f(1)$, which also takes one of those two values.

Readers will not find it difficult to see that there are only four functions with these characteristics – the one that always has the value 0; the constant function equal to 1; the one that on 0 takes value 0, and on 1, value 1; and the one that associates number 1 to 0 and number 0 to 1. As the possibilities are finite, they are all computable functions, as it is possible – at least in theory – to describe *ad hoc* a set of instructions that calculates their values. However, a description of how to calculate the image of some of the values could be so complex that we would not be able explicitly to build any Turing machine to compute the function. Let's look at an example that was suggested by Arturo Sangalli.

Let's consider the function defined on the numbers from 1 to 9 which makes value 1 correspond to n if the decimal expansion of number π contains n consecutive figures n (for example, the combination 4444 if $n = 4$) and otherwise, 0. With this definition, the value of $f(1)$ is 1 because in the decimal expansion of π, which begins with 3.141592... a 1 appears (for example, in the first position).

Analogously, $f(2)$ is also equal to 1, but to find the chain 22 it is necessary to go through the first 135 decimals: ...4460955058 22 31725359408... The following table has been drawn up by means of a program which readers can use to carry out other experiments, if they wish, at: http://www.angio.net/pi/bigpi.cgi.

Pattern	Position	Pattern	Position
333	1,698	666666	252,499
4444	54,525	7777777	3,346,228
55555	24,466	88888888	46,663,520

It can be deduced from the table that our function takes the value 1 in the first eight natural numbers, as π contains the patterns 333, 4444, 55555, 666666,

7777777 and 88888888. To calculate the value of $f(9)$, we could imagine a program that would run through, one by one, all the figures of π until it found the required combination, that is, a block of nine consecutive 9s. If it does in fact exist, sooner or later the program will find it, then the *output* will be 1. It doesn't matter how long it takes because, as we have said several times, we are dealing here with an ideal machine, without the physical limitations that computers have. However, if there were not a block formed by nine consecutive 9s, the program would never stop, and we would not be able to decide the value of $f(9)$. This approach, therefore, will never enable us to know if f is computable, unless we can prove first that somewhere in the decimal expansion of π there are nine 9s in a row. But in that case the program would be pointless, as the same argument would prove that the value of $f(9)$ is 1. Although the most evident approach does not provide results, this function is computable. To prove it, it is necessary to reason as we did before – since the number of functions defined from 1 to 9 that take values 0 and 1 is finite (in this case, 512, somewhat less manageable than our four friendly functions defined in 0 and in 1 with values 0 or 1), there is a Turing machine that computes each of them. Here we have an example of a computable function that the Turing machine cannot explicitly describe.

Another kind of computable function is the recursive function, in which $f(n)$ can be calculated from the values taken by the function in other numbers smaller than n. A large proportion of the functions that mathematicians use on a daily basis are recursive, but are all the functions computable? Alan Turing immediately came to the conclusion that the answer is in the negative: there

IF EVERYTHING WERE JUST A NUMBER?

In his tale *The Library of Babel,* Argentinian writer Jorge Luis Borges suggests that all the information in the Universe could be stored in a single book, which "would consist of an infinite number of infinitely fine pages". But why store it in such an unwieldy fashion if it could all fit in one number? One of the most mysterious conjectures that mathematicians are facing these days is to prove that within the decimal figures of the number π, the ratio between the length of circumferences and their diameters, any number pattern we can think of appears sooner or later. If that is so, not only would the combination 999999999 appear sooner or later, but any message, past, present or future, can be coded within its decimals.

are a lot of functions that no Turing machine can calculate and, still worse, if a function is chosen at random, it will almost certainly not be computable. At the same time, on the other side of the Atlantic, the logician Alonzo Church (1903-1995) was coming to exactly the same conclusions at Princeton University by means of the development of a formal system, which was given the name 'lambda calculus'. Both ideas were so innovative that the only person the publishers of the *Proceedings of the London Mathematical Society* could find to assess Turing's article was Church himself. Thus began a period of fruitful cooperation, and though it was interrupted by the war it would enable the scientists to formulate the principle that is today known as the 'Church-Turing thesis'. There might be other definitions of computable function, but if the thesis is accepted, all of them would be equivalent to the existence of a Turing machine that calculates the values of the function.

To prove that hardly any function is computable, Alan Turing made use of an ingenious variant of Cantor's diagonal argument, which we studied in detail in Chapter 2. There, we saw that there was no way to arrange the sequences of 0s and 1s in a list. As soon as we supposed that one could be placed after the other, by modifying the values of the elements of the diagonal, we managed to construct a sequence which, despite being formed exclusively by 0s and 1s, did not coincide

Alonzo Church, who introduced lambda calculus, and collaborated with Turing.

with any of those on the list. This same reasoning enables us to conclude that the functions are uncountable.

After all, what is a function? We said that it is a method to transform 0 into $f(0)$, 1 into $f(1)$, 2 into $f(2)$, and so on into the infinite. Therefore, all the information of f is contained in the sequence of numbers $f(0), f(1), f(2), f(3)\ldots$ To simplify things, think of functions that only take the values 0 and 1; for example, the function f whose value is 0 when a number is even and 1 when it is odd. In this case, all the information of f appears in the sequence 0 1 0 1 0 1 0 1 0 1..., as, if we want to calculate the image of n we just have to advance to the nth position and see what symbol we find. We hope we have convinced the reader that the functions that only take values 0 and 1 are exactly the same as the infinite sequences of 0s and 1s. Therefore, they are not enumerable!

As each Turing machine computes one single function, the first thing that has to be proven, so that the hope that all the functions are computable continues to make sense, is that there are at least as many machines as functions required to be calculated. But that is not the case: Turing proved that the infinity of his machines was much smaller. In order to see that the functions are not enumerable, firstly it was necessary to codify them by means of strings of 0s and 1s. In the case of Turing machines, we already have a method of symbols to write them, as any machine is simply a finite list of instructions, and each of them can be translated into a few symbols. As we saw, (#1, 1, L, #3) says the same as 'Instruction number 1: if the symbol seen is 1, move to the left and go to the third instruction'. Once we have expressed the Turing machine as a sequence of instructions written in this way, the reader can look for some type of order that enables all the Turing machines to be written in a list.

However, for what follows it is better for us to make use of the same *gödelisation* process that we studied in Chapter 4. Let's remember that this was a way of assigning some gigantic natural numbers to each formula of first order logic. The formula can be reconstructed from the number. This procedure, applied this time to the code of the Turing machines, enables us to encapsulate all the information on the program in one single number. As happened with *gödelisation*, not all the numbers represent a Turing machine, but only those numbers that fulfil certain properties. On the one hand, there are infinite Turing machines, but, on the other, that infinity cannot exceed that of the natural numbers, as every Turing machine is codified using one of them. Thus we have proven that Turing machines are countable. And, therefore, so are computable functions.

The halting problem

Leibniz's dream of building a machine capable of distinguishing true statements from false ones was taken up by David Hilbert in the 20th century. As we pointed out in Chapter 3, Hilbert's programme for eradicating paradoxes from mathematics did not only consist of providing mathematics with the most solid foundations possible. The ancients had already done that, beginning with Euclid, and it had not worked. To be absolutely certain that in the future no other Russell would pull another paradox out of his hat, the mathematical task of cementing the edifice of logic had to be accompanied by a metamathematical 'structural engineering' which would prove that the pillars really could support the weight of the roof. The first two questions that Hilbert put to himself were whether mathematics was both complete and consistent, in other words, if the true coincided with the provable, and whether there was not a risk of encountering contradictions. Three years before Gödel proved that in the case of arithmetic these two requirements were incompatible, David Hilbert and his assistant Wilhelm Ackermann (1896-1962) added a third question, which they revealed at the plenary conference at the International Congress of Mathematicians in 1928.

The decision problem (*Entscheidungsproblem*) questioned whether there existed any algorithm capable of receiving a mathematical statement as 'input' and returning it as true or false 'output'. While it was reasonable to demand that axioms were recursive, it was not the same with the set of theorems, as we shall see very soon. But first let's recreate a scene that is related to Hilbert's new problem, a scene that the author witnessed at another International Congress of Mathematicians, this one held in Madrid in August 2006.

By the exit of one of the conference halls, a certain mathematician was holding a conversation with someone he had mistaken for a journalist. After an exchange of jokes about a gang of thieves who had robbed some of the attendees by pretending to be police officers, this second person enquired as to what line of business the first person was in. There are few more dangerous questions than that one, the likely outcome being a monologue lasting half an hour in which the speaker's enthusiasm increases at the same rate that the listener's decreases. But this time the mathematician had decided that the journalist would not be able to understand anything, so he just explained, "Well, you see, I have a little machine I put statements in and it tells me if they're true or false", to which the supposed journalist, who up to that moment had not let on, replied, "Brilliant! Let's see if one of these days you can lend me your little machine, because I work with loads of mathematical conjectures and I haven't the faintest idea whether they're true or not."

We would all love to have that program the mathematician was boasting about to his mistaken journalist, but with his research into computable functions, Alan Turing demonstrated that it was impossible. To do so, he imagined a universal machine that not only admitted numbers as input but also the instructions of any Turing machine. If the instructions corresponded to what we now call a program, the universal machine was the computer itself, capable of imitating, at least at a theoretical level, what any Turing machine does. This imaginary computer foresaw the architecture of our computers which would arrive several years later. The editors of *Time* magazine were not exaggerating when, on electing Alan Turing among the most important people of the millennium, they suggested we should always remember that: "Everyone who taps at a keyboard, opens a spreadsheet or a word-processing program, is working on an incarnation of a Turing machine". By making use of this *avant la lettre* computer, Turing showed the absurd result to which the existence of such a 'truth machine' led.

Let's look at Turing's solution to *Entscheidungsproblem*. To begin with, our hero supposed that Hilbert's dream was obtainable, that is, that there was a mechanical procedure able to decide in finite time if a statement was true or false. In particular, such an algorithm would allow us to know if the assertion 'The Turing machine T halts when it receives input n' is true or false. As we have already shown, thanks to the *gödelisation* method we can associate a number to each Turing machine, so that all its structure is codified within it. When n is the number of a Turing machine, we shall denote it by $T(n)$. With this notation, the problem we want to solve can be stated in the following terms: does the Turing machine $T(n)$ halt when it is fed as input number m? We must emphasise the fact that, if the perfect machine that Hilbert imagined did exist, then it would not only be able to answer this question in some cases, but always, whatever the values of parameters m and n. It is, therefore, a function of two variables that takes the pair of numbers (m, n) and which predicts if the Turing machine associated with n halts when it is started up with a tape representing number m. In the example of number π, if we give the name t to the number of the Turing machine that runs through the decimal figures searching for the required combinations, when parameters $(9, t)$ are entered, our function will reply 1 if among the digits of π nine consecutive 9s appear (because then the machine halts), and 0 if not (as it will carry on running indefinitely).

However, by supposing that there exists a Turing machine H (H for halt) that solves this problem, we immediately arrive at a contradiction. To see this clearly,

it is worth repeating once more what it is that H does. It is a Turing machine whose inputs are pairs of numbers (m, n) and whose outputs only take two values: 1 if the Turing machine $T(n)$ halts for the initial value m, and 0 otherwise. In other words, if no Turing machine exists represented by the number n (as not all natural numbers are numbers of a Turing machine) or, even if the Turing machine exists, the program continues running indefinitely when parameter m is entered. This is one of the bugbears of IT specialists but, undeterred, they gave it the sexy name 'infinite loop'. What matters to us here is that, if we had this machine, we could easily construct another Turing machine, which we would call C (C for contradiction), whose input would be a single number n and which would operate in the following way:

– If the Turing machine $T(n)$ halts when input n is entered (in other words, if the value of $H(n, n)$ is 1), then C never halts.

– If the Turing machine $T(n)$ continues running indefinitely, starting from value n (that is, if the value of $H(n, n)$ is 0, then C stops as soon as it begins.

In Chapter 2 we saw how the liar paradox, which so tormented dear old Epimenides, emerged when a Cretan was made to say that all Cretans were liars, or, alternatively, when a sentence said of itself 'This statement is false'. Later, we showed how Gödel made use of self-reference to construct a true, but unprovable, statement – 'This sentence is unprovable'. After this reminder, the reader will surely know how to finish off the argument. We have defined a Turing machine C, which halts or continues running non-stop depending on what another machine $T(n)$ does. But, what happens if, for input of C we enter C itself, in other words, the associated number c? If the machine $T(c)$ halts, then C does not halt. If, on the other hand, $T(c)$ goes into an infinite loop, then C halts. But C and $T(c)$ are the same machine! They cannot do different things! Supposing that the halting problem had a solution for any value of m and n, then we have reached a contradiction, as when the fiendish self-reference whispers in our ear 'choose c!' it turns out that the same machine behaves in two different ways.

The dream of Hilbert and Leibniz was a utopia. The same game of mirrors, had first prompted Bertrand Russell to embark on the task of rebuilding mathematics on more solid foundations. Afterwards that enabled Gödel to prove that the optimism of his era was not justified, and now Turing was using it again in his solution to

Entscheidungsproblem, this time in the guise of theoretical machines from which computers would be born.

We said that logic does not deal with how we reason in everyday life, but with how we ought to reason to be sure that the conclusion we reach is true. And, indeed, up to now we have only looked at formulae in which the values of 0 and 1 appeared empty of meaning. Black or white. In the next chapter, however, we shall try to describe what a world would be like with greys; it would be less grey, but also more insecure...

Chapter 6

All's Well That Never Ends

To learn how to gauge milligrams requires a long and difficult apprenticeship that only the purest persons have the courage to attempt.
Marguerite Duras, French writer and film director

He knew what he was doing when he took her to that Japanese restaurant. Although he was in no doubt about his charm, if his seduction techniques did not work with the stories of his many travels, he could still save the night by skilful handling of his chopsticks and woo her by ordering dishes with sensuous names. When the waitress came to take their orders for dessert, everything was going well. It was clear that the waitress had learned their language as a child, so when she turned back to ask if they wanted the tea truffles they had just ordered "with cream, without cream, or what?", the couple realised that this was not a mistranslation. The man who had been bragging all evening took a chance with: "With the what!" The waitress returned, smiling, with a plate of truffles with which there was just a little cream on the side of the plate. It was then that the couple looked into each other's eyes and said together: "She has no principle of non-contradiction."

Fuzzy logic

Despite their varied appearances, all the sets we have looked at up to now share a common property: given any element and any set, the question 'Does the element belong to the set?' only has one answer, yes or no. It is true that the description that defines it could be so complicated that we wouldn't be able to answer, but the important thing is that in some corner of the mathematical world an answer – yes or no – is written. That is what happened in the example of the numbers that have decimal expansions that contain every pattern that can be imagined, as we showed in the previous chapter. We don't know if π falls within the set, but only one answer is valid. The propositions of logic fit into this pattern. They are true or false, and any other possibility is excluded. So much so, in fact, that the

two main paradoxes that we have dealt with (Russell's and the liar one) arise when it is not possible, not even from a theoretical point of view, to reply true or false, whether it is a member or it isn't a member. And it isn't that our rule admits an exception, but instead says that neither the set of all the sets that are not members of themselves, nor the sentence 'This statement is false' are formally correct. This is because the membership relationship is only applicable between objects of a successive type or because the concept of truth does not belong to language but to metalanguage. In some ways, set theory and logic are cliffs. True is right on the cliff edge, and a breath of wind is enough to send it into free fall towards false below. However, in actual geography, what is more abundant than cliffs are those gentle slopes that very gradually fall down to the sea.

A few years ago, the parlour game Scattergories became all the rage in some countries thanks to a TV advertising campaign. It's a game in which a letter from the alphabet is randomly chosen and then players have to write words belonging to different semantic fields that begin with it. For example, we are given the list 'Sports. Song titles. Parts of the body. Ethnic food. Things that are shouted'. On rolling the icosahedral die, the letter T appears, a possible answer would be: "Tennis. *Take Me To The River*. Tonsil. Tajine. Taxi!" In the TV ad, a young man is seen angrily leaving a house taking the Scattergories game with him because his friends had not accepted his answer 'boat' for the category of 'aquatic animals'. The friends finally decide to give in, seeing as it's the only way they can continue playing, but the young man gets back to his old tricks in the next game, this time with the letter O. His friends are left wondering: 'Should we accept octopus as a pet?' If there are living things with classifications that can raise serious doubts, the set of pet animals is even more poorly demarcated. No one would doubt that dogs and cats form part of it, and with the same certainty we can say that neither wolves nor elephants belong to it.

However, while some folk would include tarantulas in the 'animals I don't want anywhere near me' category, there are those who find it entertaining to feed them crickets through a hatch in their cages. No better defined than pets are other sets that we encounter in daily life, such as the set of good-looking people, of good restaurants or of funny jokes. The first person to propose a theory to respond to these situations was the Polish logician Jan Łukasiewicz (1878-1956), who in 1917 introduced three-valued logic, in which propositions could be 'possible' as well as true or false. For example, a person 1.5 metres tall is short; one of 2 metres is tall,

THE REVENGE OF THE LIAR

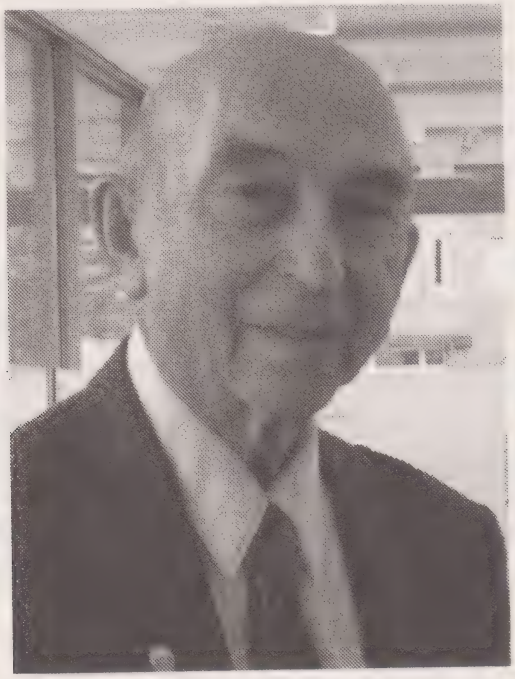
If we take another look at the liar paradox in the light of the three-valued logic of Łukasiewicz, we realise that the contradiction has disappeared. The essential component of our analysis consisted of deducing that if a statement of the type 'This sentence is false' was not true, then it must be false. However, in the new logic, there are statements that are neither true nor false, but possible. As the fundamental reason behind the paradox is deeper than that principle of bivalence, it is possible to modify it so that it is still valid in three-valued logic. Let's take the statement 'This sentence is not true'. Affirmations can be divided into three classes (true, false and possible), so we'll reason case by case. If the statement is true then what it says must be right, so it is not true. If, on the other hand, the statement is false or possible, then it is not true, but as that is precisely its contents, it must be. Neither is it possible in the new logic to attribute a value of true to the statement 'This sentence is not true'.

while one of 1.75 metres is 'possibly tall' or 'possibly short', depending on whether we compare them with, say, a stable full of jockeys or a team of basketball players.

Including 'possible' among the values of truth is a step forward with respect to the black-and-white world of classical logic, but even so it is insufficient as it refers to an indeterminate point, and what we are interested in is being able to make decisions. Let's suppose that a journalist is wondering whether to resign after a change in his newspaper's editorial policy. We'll use *P* to denote the proposition 'I do not agree with the newspaper's new ideology'. So, the structure of a classical decision would be 'If *P* is true, I resign' and 'If *P* is false, I stay'. As being in agreement is always a question of nuances, the journalist could do with a third value of truth. But how do we interpret 'possible'? If *P* is possible, do I resign or do I stay? What barrier separates one reaction from the other? If we want a logic that helps us to make this type of decisions, we have to be much more precise.

It is here that the Berkeley professor Lotfi Zadeh appears on the scene. In 1965, Zadeh proposed that the membership of a set or the truth of a proposition should take any value between 0 and 1. Using this method, players of Scattergories could rule that the only valid answers are those that belong, for example, more than 0.6 to the semantic field in question, and the journalist could decide that if his disagreement with the paper's new editorial policy exceeded, say, 0.45, then he would resign. Zadeh christened the new sets with the term 'fuzzy', which can mean hazy or diffuse, that

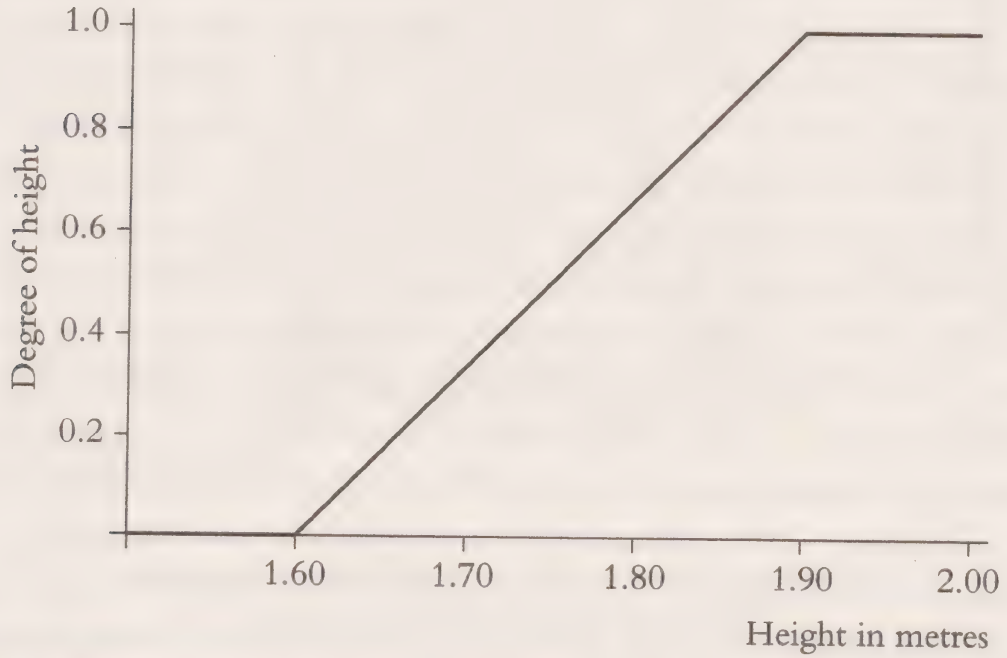
which lacks well-defined limits. In fuzzy sets, therefore, the question 'Is the element a member of the set?' has an infinite number of answers.



Lotfi Zadeh, the creator of fuzzy logic (source: Wolfgang Hunsche).

The reader will perhaps be tempted to interpret fuzzy sets in terms of probability. The explanation may seem clearer that way, but to say that the membership of an element to a set is the probability that it is contained within it would be to betray the spirit in which Zadeh invented fuzzy logic. Let's look at what happens when a coin is tossed. Ever since we were little children we have known that the probability that it comes up heads is 50 per cent, and that means that if we toss the coin a large number of times, say ten thousand, more or less half of the results will be heads, the other half tails. However, each toss only gives one result: heads or tails, yes or no, a member or not a member. Probability, at least in its simplest version, reflects our limited knowledge of things. If we knew exactly the force used to toss the coin, if we could become the god Aeolus and control the winds, then we would be able to predict the result. This means that the underlying principle to this simple incarnation of the theory of probability basically coincides with classical logic. In the fuzzy world, when the coin is tossed the result may be only heads, more heads than tails, more tails than heads, only tails, or any of the intermediate variants, expressed with infinite precision.

Unlike classical sets, whose border is an abyss, the sets studied in fuzzy logic are delimited by a membership function shaped like the slope of high land gradually flowing down into the sea. Let's take, for example, tall persons. If we decide that 1.60 metres is completely short, and that from 1.90 upwards is completely tall, then the set membership function looks like this:

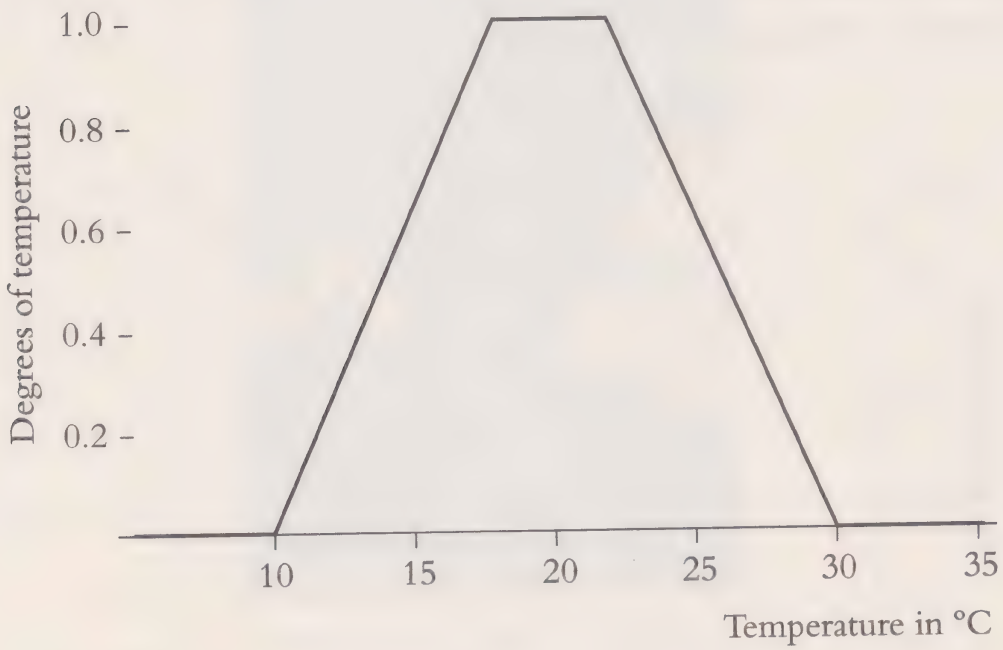


The membership of a fuzzy set for tall people. The function takes the shape of a slope.

By making some calculations, it can be shown that everyone who is less than 1.60 metres tall will have a degree of membership to the set of tall persons of 0. If their height is greater or equal to 1.90, they will be completely tall, and if their height is, like most of us, between those values, then to calculate the value of membership their height (measured in metres) must be multiplied by ten, 16 subtracted from it, and the result divided by three. Just knowing that a person's height degree is 0.5, which is the case of the author of this book, is enough to discover how tall they are.

In other cases, the membership functions may take the shape of a triangle or trapezoid. If we consider, for example, that at a temperature below 10° C it is very cold, above 30° C it's too hot, and that a temperature that is perfect is between 18° C and 22° C, then the membership function for pleasant temperatures would

look like the graph shown below. Comparing it with climate data would not be a bad way of choosing where to live! Or at least for ruling out some places...



Membership function for the fuzzy set of pleasant temperatures. The function takes the shape of a trapezium.

By adapting to the scale of greys existing in reality, fuzzy sets allow us to solve some paradoxes, taking the term in its widest meaning. Let's imagine that in a bar we are served a really bitter cup of coffee. Leaving aside exceptions, to drink it we will have to add a little sugar. The reader will agree with us that, if we add just one grain of sugar, the taste of the coffee will not change one bit; therefore, we can conclude that the operation 'adding a grain of sugar' does not change the bitterness of the coffee. Let's put in another grain, then another, after that, still another, until we get up to ten sachets of sugar. If our principle is correct, as none of the steps changes the taste, the cup of coffee with ten sachets of sugar will continue to be as bitter as the one we were originally given, which is, to put it mildly, worrying... As it is easy to see, what happens is that sweet coffee is not a set in the classical sense like those that we studied in the previous chapter. There is not an abyss, but a continuous slope between being so bitter that it cannot be drunk and being so sweet that it's sickly. Though our palate is not sufficiently sensitive to notice the change, when a grain of sugar is added to the cup its membership degree of the set of sweet coffee increases

a little, however small the change may be. If we add a little more, it continues to grow, and that's how, after ten sachets of sugar, the coffee ends up becoming so sweet that it's undrinkable.

When generalising a concept in mathematics, as Zadeh was attempting to do by introducing fuzzy logic, it is essential to be sure that the new formalism is still valid for studying the initial objects. Classical sets are a very particular example of fuzzy sets – precisely those in which, among the infinite possibilities that it has available, the degree of membership only takes values of 0 and 1. However, it is not so clear how to generalise the fact that a set is contained in another, or the operations of union and intersection, which, as we saw back in Chapter 3, were fundamental in set theory. These were some of the questions that Zadeh answered in his article in 1965.

In what follows, A and B will be two fuzzy sets, whose associated membership functions we shall denote by f_A and f_B. This only means that, given an element x, the number $f_A(x)$, which indicates the membership degree of x to the set A, is between 0 and 1, and that the same thing occurs with $f_B(x)$. By using this notation, Zadeh stipulates that A is contained in B if, whatever the element x is, the number $f_A(x)$ is smaller than or equal to $f_B(x)$. Let's look at an example. Instead of considering that up to 1.60 m is completely short, and that from 1.90 is completely tall, we could reduce the limit a little, so that persons 1.50 m tall are considered completely short, and from there upwards, the membership degree increased up to 1.90 m, like before. In this way we would obtain another fuzzy set of tall people. In this one, the author's membership degree is now not 0.5, but 0.625. Well, then, what Zadeh tells us is that the first set that we had defined is contained in this latter one, which fits in with the intuitive idea that all people that were tall in accordance with the strict yardstick are still tall once the limits have been lowered.

Lotfi Zadeh had studied electronic engineering and by inventing fuzzy logic he suspected that its application could be used in IT processing and in pattern recognition, two fields where imprecision is the predominant feature. The course of history has shown that Zadeh underestimated his idea, and no country was to highlight that more than the one whose inhabitants eat their tea truffles with cream, without cream, or what. By the end of the 1990s, in Japanese shopping centres they had begun selling fuzzy photocopiers and washing-machines, skyscrapers in Tokyo were not considered up to date if they did not have a fuzzy lift which reduced waiting time to a minimum. As the commercials for these washing-machines proclaimed: the fuzzy age has begun!

FUZZY WASHING-MACHINES

To shorten the time it takes to do the washing and improve the quality of the wash, we would find it useful to use exact quantification on whether the clothes are very dirty, a bit dirty or practically clean. In their simplest versions, fuzzy washing-machines assign a value of dirtiness of between 0 and 1 for each article. Then, onto a fixed basis of ten minutes' washing, more time is added, depending on the degree of dirtiness of the clothes. The machine might consider, for example, that the clean clothes (0) have now been washed for the basic time period and that for each very dirty article (1), two minutes more are needed. So a shirt that is half-clean/half-dirty would mean an increase of a minute in washing time. Other more sophisticated models also include sensors for the amount of grease, which is more difficult to eliminate than other stains, or of the load in the machine, all with the aim of saving energy.

Complexity

'Love' and 'justice' are concepts that are too subtle to be governed by yes or no logic. The grey zone that opens up between 'He loves me' and 'He love me not', between guilt and innocence, is the domain of fuzzy logic. As complexity increases, a need for new thinking arises. It would be useful, therefore, to have estimations of how difficult the concepts are, but 'complexity' is one of those notions that defy any definition. Not even in the mathematics world can easy problems be precisely distinguished from difficult ones. In the case of Turing machines, while in the previous chapter working with ideal computers allowed us to obtain theoretical results about the problems that a machine is not able to solve, what now interests us is to determine what calculations can be carried out bearing in mind real computers' limitations on memory and running time. That is what will determine, while we await a better definition, whether a problem is easy or difficult.

Firstly, we could establish that the complexity of a task is the number of operations necessary to carry it out. Let's imagine a businessman who must visit a number of cities and then come back to the initial starting point. His objective is, of course, to cover the minimum possible distance. If those cities were, for example, Paris (P), London (L), Berlin (B) and Rome (R), and the executive started off in Paris, then his secretary could organise his agenda in six different ways: PLBRP, PLRBP, PBLRP, PBRLP, PRBLP and PRLBP. Bearing in mind the approximate

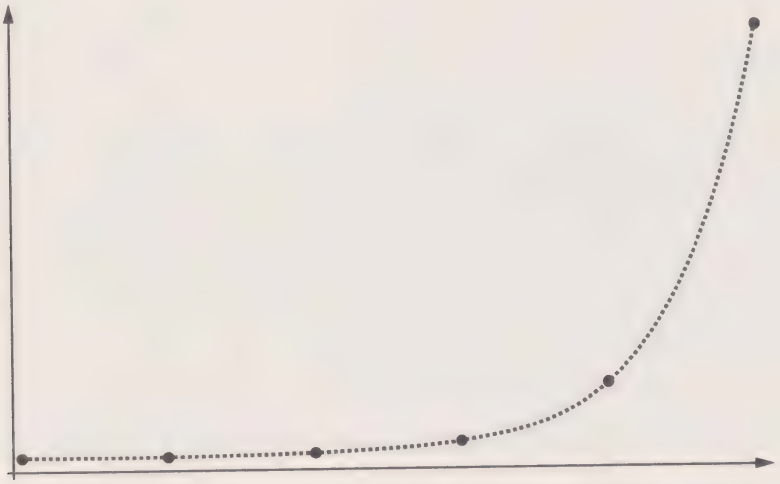
distances: Paris-London (455 km), Paris-Berlin (1,050 km), Paris-Rome (1,435 km), London-Berlin (1,095 km), London-Rome (1,855 km) and Berlin-Rome (1,515 km), the secretary could calculate the number of kilometres for each route and choose the shortest:

Route	Km	Route	Km
PLBRP	4,500	PBRLP	4,875
PLRBP	4,875	PRBLP	4,500
PBLRP	5,435	PRLBP	5,435

The table shows us that the optimal sequence of journeys is Paris-London-Berlin-Rome-Paris or, alternatively, starting from the other end, Paris-Rome-Berlin-London-Paris. To solve this problem, the six alternatives have been analysed one by one. But what would happen if, instead of three cities, we had to visit four, five or any number? Just with 20 cities, the average computer would take some eighty thousand years to find the fastest way to travel. The time lost by our businessman through making the wrong decision simply pales into insignificance. By trying to find the solution by 'brute force', there are now not six cases to consider but the number obtained from multiplying $1 \cdot 2 \cdot 3 \cdot \ldots$ and so on up to 20, which has 19 digits. It is what we mathematicians call the factorial of 20 and which we show by putting an exclamation mark after the number. Thus, $3! = 1 \cdot 2 \cdot 3$ is 6; $4! = 1 \cdot 2 \cdot 3 \cdot 4$ comes to 24; and in general, $n!$ is the product of the n first natural numbers.

The factorial is an example of a function that is very simple to calculate from a theoretical point of view, but which in practice drives computers to the limit. As we commented on in the previous chapter, all recursive functions are computable. Let's remember that a function is recursive if $f(n)$ can be calculated from the values taken by the function in other numbers smaller than n. Well, then, the factorial is the typical case of a recursive function, as, if we want to calculate $4! = 1 \cdot 2 \cdot 3 \cdot 4$, we can first do the product $1 \cdot 2 \cdot 3$ and then multiply by 4. But, what is the product $1 \cdot 2 \cdot 3$? Precisely 3!, so, by knowing the value of 3!, just one operation is enough to obtain the factorial of 4. In general, $n! = (n-1)! \cdot n$, and that proves that the factorial is a recursive function and, therefore, computable. For a Turing machine with all eternity ahead of it, the calculation of $n!$ is no mystery. However, in practice the

values of the function grow so quickly that they soon become unmanageable, as can be seen in the following graph:



A graph showing the growth of the factorial function.

The previous example would be no more than a curiosity if it were not that the factorial counts the number of permutations of the finite sets, in other words, how many different ways their elements can be arranged. So, the statements '3! = 6' and 'the set {1, 2, 3} can be written in six different ways (123, 132, 213, 231, 312 and 321)' express the same information. As the naive analysis of many problems similar to that of the jet-setting executive requires a one-by-one examination of all the permutations of a set that could be formed by many elements, the speed with which the factorial grows has tricky consequences for IT. It is one of the first things that makes a problem 'difficult'. The 'easy' ones, on the other hand, will be those that are not only solvable from a theoretical point of view, but also in practice, in reasonable time. They are often denoted by the letter P (P for polynomial), because the number of operations necessary increases in accordance with the size of the data more or less at the same rate as polynomials do.

What IT specialists notice is that there are problems for which it is very difficult to find the solution, while proving that it is the correct result turns out to be easy. Let's return to our hotel from Chapter 2, which this time will have a finite number of rooms. Let's suppose that a group of four hundred people want to spend the night there on some dates when there are only one hundred vacant rooms. Choosing them without having to follow any criterion would be very easy, but the booking form

THE INVENTOR OF CHESS

According to legend, a Persian king wanted to reward the inventor of chess by offering him whatever he wanted, however much it cost. The wise gamer surprised him with a request that at first glance seemed to be very humble: he wanted a grain of wheat for the first square of the chess board, two for the second, four for the third, and so on, each time doubling the amount for the previous square up to the last. Annoyed at what seemed to be mockery of his generosity, the king ordered the wish to be immediately carried out and the grain dispatched to the inventor of chess. How surprised he must have been the next day when his advisers informed him that there was not enough grain in all the barns in the world to fulfil the wise man's request. The function that began by taking the values 1, 2, 4, 8... grew so fast that 18,446,744,073,709,551,615 grains of wheat were needed.

had attached to it a strange request. There are a number of people that get on so badly with each other that under no circumstances could they sleep in contiguous rooms. It is unthinkable to try and solve the problem by examining all the possible choices, one by one, of a hundred people among four hundred, and yet, once a solution has been proposed, it is enough just to check that two incompatible people do not appear together in the list of allocated rooms. The receptionist could do it in

just a few hours without the need of a computer. We mathematicians refer to these problems – ones that are difficult to solve but easy to prove – as NP.

Up to now, we have referred to the complexity of problems as if it was a property that was intrinsic to the statement. This viewpoint is innately erroneous, because what is easy and what is difficult is not the problem in itself, but our way of resolving it. It might be that we have found a solution that requires a great number of operations, but that another simpler one exists. In this case, our solution will be in NP, while the problem would belong to P. The strategy for the businessman to optimise his business trips consisted of examining all the possible routes one by one, without doing any reasoning. The table, however, shows that on reversing the order of the itinerary, the distance does not vary. It makes no difference choosing Paris-London-Berlin-Rome-Paris or Paris-Rome-Berlin-London-Paris, so it would suffice to deal with half the cases. In practice, this simplification does not improve things as half an enormous number is still enormous. Rather, its significance is more strategic. If in the first solution we missed such a trivial detail, how many other tricks of the same kind will we have overlooked? We said that the viewpoint was, a priori, erroneous, because the truth is that we do not know if there are difficult problems in the absolute sense. Although the executive problem is one of the candidates, no one has been able to prove yet that all its solutions are difficult.

Another objection that this concept of complexity raises is that it does not help to distinguish between tasks that involve the same quantity of operations. According to our definition, memorising a password of 12 symbols is easy or difficult whatever the

P VERSUS NP

As we saw in Chapter 3, the symbolic beginning of 20th-century mathematics had taken place in August 1900 in Paris with Hilbert's list of the 23 problems. Also in Paris, but one hundred years later, a commission of experts from the Clay Mathematics Institute met to choose the seven questions still open that would, in their opinion, mark the path of research in the coming century. The fourth problem on the list, known as 'P versus NP', is to find out if there are NP problems in the absolute sense or if, on the contrary, any problem whose solutions can be verified in polynomial time can also be solved quickly once an algorithm ingenious enough has been found. There's a million dollars waiting for whoever can answer the question. So, after all, maybe mathematics does pay.

characters are, as it will always need 12 operations: memorising the first, memorising the second and so on up to the twelfth. However, nobody in their right mind would think that memorising the passwords 111111111111 and 6u0yfz3eq85s require the same effort. While the first one can be compressed into 'twelve 1s,' the only way to describe the second is character by character. It was with this example in mind that in the mid-1960s the Russian mathematician Andrei Kolmogorov proposed replacing the number of operations with the number of instructions. From then on, the complexity of a chain of symbols would be the minimum length of the algorithm necessary to generate it.

Let's imagine a Turing machine whose task consists of writing a certain chain of 0s and 1s, which we'll call *s*. As we saw in the previous chapter, the machine will have to be given a series of instructions of the type 'If the symbol seen is 1, move to the right and go to order #2'. In this simplified version, we'll say that the complexity of *s* is a natural number *n* if there is a Turing machine described by means of *n* instructions whose *output* is *s*, and if no machine with fewer orders can generate our sequence. We thus get a function *K* (K from Kolmogorov) that associates its complexity to each chain of 0s and 1s. Let's take, for example, the succession 1111... The reader can see that, if, as *input* a tape full of 0s is entered into a Turing machine whose only instruction is 'Instruction #1: If the symbol seen is 0, write 1 and go to instruction #1. If the symbol seen is 1, move to the right and go to instruction #1', then the result we get will be the succession 1111... This means that its complexity is the minimum possible, $K(s) = 1$, because one single instruction suffices.

One surprising consequence of the new concept of complexity is that computers are not able to generate random infinite chains of 0s and 1s. In the intuitive sense, a succession is random when, on account of its internal structure, it is impossible to predict what the next term will be. That means that a random sequence cannot be compressed into a description shorter than itself; in other words, its complexity is infinite. However, all IT programs work with a finite number of instructions (remember the definition of Turing machine given in the previous chapter). Therefore, the chains of 0s and 1s that they generate, however devilish they may seem, will always be of a finite complexity. Computers can only write *pseudorandom* sequences and that is why many physicists have been trying for years to make use of the indeterminacy properties of atoms to build truly random sequences.

At the same time, Kolmogorov complexity has many similarities with the Oxford librarian's paradox that we explained at the end of Chapter 2, and which

consisted of studying the natural numbers that can be described in 15 words. As there is only a finite number of expressions with 15 words, this set is also finite. Therefore, from among all the numbers that are not members of it, there will be one that is the smallest. Let's call it n. So, n is 'the smallest number that we cannot describe in fewer than 15 words', but this description has 12 words! It is quite natural to ask ourselves whether the definition that we have just introduced will not lead us into contradictions, and the answer is another surprise: if the function K were computable, in other words if there was a Turing machine which, on being fed the *input* of a chain s of 0s and 1s, returned as its *output* complexity $K(s)$, then a reasoning similar to the halting problem would enable us to reproduce the librarian's paradox in the formal language of arithmetic. Therefore, the only possible answer is that the complexity is not computable, and that is enough to solve the librarian's paradox that was still outstanding. The fact is that the expression 'describe in 15 words' is not correct because it belongs not to language but to metalanguage.

Gödel, Turing and artificial language

On earlier pages we have been content to discuss the concept of complexity only in the field of maths, where the reader will have been able to note that there are numerous difficulties. Our initial purpose was even more ambitious: we wanted to know what it means when we say that the ideas of 'love' and 'justice' are complex. Little by little, mathematics have served as inspiration behind another theory of complexity, which could be summed up by the saying 'The whole is greater than the sum of its parts'. The words 'lucidity', 'wound', 'Sun' and 'closest' each have a very precise meaning. We could consult the definitions in several dictionary and learn about their etymology. However, when the French poet René Clar writes that "Lucidity is the wound closest to the Sun", from four words that we knew perfectly well something has emerged that was not in them before. The verse is greater than the sum of its parts; that's why some poetry is difficult to understand.

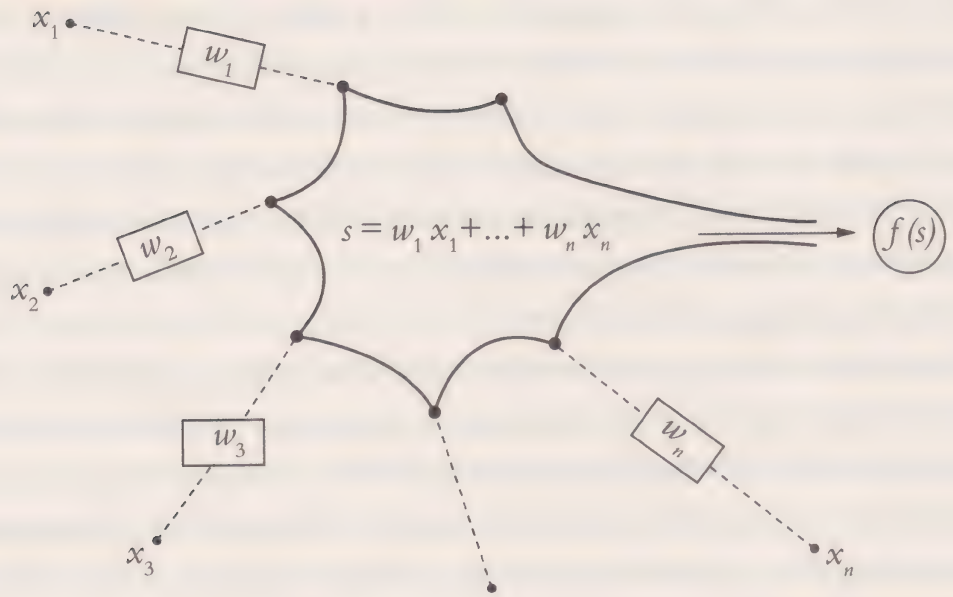
Far from being a phenomenon exclusive to language, this principle of emergence is found in an incredible variety of forms. It's present in what are called social insects, it explains the success of the Internet and is one of the keys to the study of the human nervous systems. Let's consider, for example, the humble ant, which complies the best it can with its impulse to search for food. We would never be able

to understand the complex organisation of an anthill, able to adapt to the most extreme situations, by seeing it simply as a sum of ants. The immune system, too, is more than just a collection of cells; the economy more than the set of shareholders; and the Internet is more than the sum of the isolated actions of each user around the world. Understanding how the complexity of the whole emerges from the relative simplicity of each component of these systems is one of science's great challenges at the beginning of this century.

Although the definition of 'complex system' as one in which the whole is greater than the sum of its parts is too imprecise, there is no doubt that a brain fits that definition very well. In this case, the individual components are the neurons, cells formed around a central body that receives and processes the impulses that come to it through a series of branches and transmits them to other neurons. The most widespread opinion among those who study the brain is that the network of connections that converts it into more than a group of isolated neurons is behind phenomena such as perception, intelligence and feelings. But what if we could translate this structure to IT? The first attempts to mathematically model neurons go back to an article in 1943 in which the neurologist Walter McCulloch and the logician Walter Pitts gave an elegant definition of a neuron as a function that received many *inputs* and produced a single *output*.

Up to now, all the functions that have been covered in this book took a single initial value and transformed it into another through a series of operations. In real life, however, few if any phenomena depend on one parameter alone. The modern theory of artificial neural networks, inspired by the ideas of Pitts and McCulloch, allows functions with many parameters to be computed by imitating the workings of the brain. Let's suppose that we want to calculate the value of a function f which depends on the numbers $x_1, x_2, \ldots, x_n$.

The idea is that the program receives them as if they were the electrical impulses that arrive at a neuron through its branches. As not all of them are necessarily of the same intensity, each number x_i will have to be accompanied by another number w_i, called weight and which measures the importance of each electronic impulse in comparison to the others. For example, if w_1 and w_n were much bigger than $w_2, w_3 \ldots, w_{n-1}$, that would mean that the parameters that really have an influence would be the first and the last. With the weights of the impulses at the ready, the artificial neuron calculates the weighted sum $s = w_1x_1 + w_2x_2 + \ldots + w_nx_n$ and evaluates the function in it, as it is shown in the illustration overleaf:



The innovation of neural networks is that the program we aim to solve our problem with is now not a fixed algorithm but an *open work* in which the weights can change. In fact, the normal procedure is to subject the neural network to a training period in which through trial and error the program 'learns' what the most appropriate weights are, in other words, which *inputs* need to be privileged in order to find the best solution. If the neural network is charged with recognising the human voice, for example, and one of the conclusions in the training period is that most of what is received by the first impulse is background noise, then the neurons will learn to attach little importance to it. Other tasks for which neural networks have shown themselves to be very effective are weather forecasts, and even for the flying executive problem. Computers that incorporate this technique and other more advanced ones can now solve the problem for two hundred cities.

Thanks to fuzzy logic and neural networks, the possibility that computers will be able to imitate many of the brain's activities is no longer just sci-fi speculation but has become the main goal in a booming field of study – artificial intelligence. For many years it was thought that a machine would never be able to play chess like grand masters. However many moves they could foresee, they would not become aware of their opponent's weak points or be able to take other psychological factors into consideration. Neither would they be much good at betting games. How is it possible to teach a computer to play poker if a bluff can contradict all the probabilities of winning?

The critics had to eat their words when in 1996 the supercomputer Deep Blue, which IBM had developed as the culmination of efforts going back to the 1950s, managed to beat Garry Kasparov in the first game of a chess match. However, despite the 100 million positions that Deep Blue analysed per second, in the next four games – played at a slower rate than usual – the Russian won all four. But just one year later the team had improved Deep Blue to such an extent that it won three of the games and managed to draw the fourth, and playing at the same speed as the professionals. The world champion had been defeated, but this hard knock did not prevent Kasparov continuing to defend the supremacy of human intelligence, curiously with the same argument that his competitors had used to program Deep Blue: "It is the synthesis, the ability to combine creativity and calculation, art and science, into a whole that is much greater than the sum of its parts".

Garry Kasparov works out his next move during the game he played against the supercomputer Deep Blue on 10 May 1997.

These advances simply revived the passionate scientific controversies that had set Kurt Gödel and Alan Turing at odds 50 years before. Using different methods, both mathematicians had come to the same conclusion on the definition of a formal system and on demonstrating undecidable problems. However, while Gödel distinguished between formalism and logic, mechanism and mind, Turing considered them to be totally synonymous. Taking this comparison to the extreme, in 1947 the English mathematician postulated that the best model of the human mind was the universal machine – able to imitate the behaviour of any program – which he himself had introduced with the aim of resolving Hilbert's decision problem. Turing believed that the question of whether computers can think could only be solved by experimental means. In an article that would make history, *Computing Machinery and Intelligence*, in 1950 Turing proposed an 'imitation game' for scientists to discover, by means of a number of written questions, if the person on the other side of the room was a human being or a computer. The idea behind the test was that if a machine behaved in all aspects like an intelligent being, then the simplest explanation was that it was an intelligent being.

Among other questions, Turing suggested that the candidate to being intelligent should be asked to write a poem or to carry out difficult numeric calculations. To begin with, correct solutions to the first question tilted the balance towards human beings, while a quick reply to the second would make us think of a computer. It could be objected that many people are not capable of writing a poem, or that, if the individual happened to be a vanguard poet, it would be difficult to distinguish it from randomly generated verses. Likewise, there are genuine 'human calculators', able to multiply and factorise very large numbers as fast as computers. In spite of such obstacles, everyone agrees that, with an unlimited number of questions at our disposal, we should be able to distinguish between a human being and a machine. Up to now, not only has no computer been able to pass Turing's test but, what's more, the same test is used in the process for recognising spam, the mass of junk email sent out by rogue computers.

In December 1969, 15 years after the death of Turing, Gödel believed he had discovered an error with serious philosophical consequences in the English logician's work. In his opinion, Turing had not taken into account that the mind is not static but in constant development. During a proof, formal systems are not subject to modifications, nor are machines during a calculation, but there is nothing to make us believe that the living mind does not change when reasoning. It can, therefore,

DIALOGUE FROM THE FILM *I, ROBOT* (ALEX PROYAS/JEFF VINTAR/ISAAC ASIMOV, 2004)

The film's protagonist, a police officer named Spooner, is investigating a murder, of which he accuses a robot called Sonny.

Spooner: Murder's a new trick for a robot. Congratulations. Respond.

Sonny: What does this action signify? (He winks.) As you entered, when you looked at the other human. What does it mean? (He winks again.)

Spooner: It a sign of trust. It's a human thing. You wouldn't understand.

Sonny: My father tried to teach me human emotions. They are… difficult.

Spooner: You mean your designer.

Sonny: Yes.

Spooner: Why'd you murder him?

Sonny: I did not murder Doctor Lanning.

Spooner: Wanna explain why you were hiding at the crime scene?

Sonny: I was frightened.

Spooner: Robots don't feel fear. They don't feel anything. They don't eat, they don't sleep...

Sonny: I do. I've even had dreams.

Spooner: Human beings have dreams. Even dogs have dreams, but not you, you are just a machine. An imitation of life. Can a robot write a symphony? Can a robot turn a canvas into a beautiful work of art?

Sonny: (With genuine interest.) Can you?

never be replaced by a computer. If you read a book that postulates against artificial intelligence it is likely that sooner or later you will find a section on the Gödelian arguments. They, however, do not refer to the objection that we have just raised but to the ideas of the Oxford philosopher John R. Lucas that the incompleteness theorems have something to say about the possibility that some day there will be truly intelligent machines. Curiously enough, Gödel never really thought too seriously that his results had any link with the structure of the human mind.

The most famous Gödelian argument against artificial intelligence is due, as we said, to the philosopher John Lucas, who studied mathematics before taking up philosophy and ancient history. In his article *Minds, Machines and Gödel*, read to the Oxford Philosophical Society in 1959, Lucas set out with rotund simplicity the reasons why the mind is not reducible to computers. As we are able to teach a machine the axioms and deductive rules of arithmetic, we could leave it constructing all the formulae of language and ask it which are true. Sooner or later, the computer would come up with the sentence "This statement is not provable" and would spend the rest of eternity trying to prove or refute it, whereas we humans see at once that the statement is undecidable due to its very contents. "So the machine will still not be an adequate model of the mind [...] which can always go one better than any formal, ossified, dead system can", concluded Lucas.

Half a century later, now hardly anyone accepts John Lucas' argument, nor the refined version proposed by the physicist Roger Penrose in 1989. What does it mean when we say that we *see* the truth of Gödel's statement? What the first incompleteness theory says is that, if arithmetic is consistent, then the proposition 'This statement is unprovable' is true, so to *see* its truth we first have to *see* the consistency of arithmetic. If we accept it as an act of faith, because we believe in a world free from contradictions, we could also program an android whose IT code included the hope that arithmetic were consistent.

This is none other than a reinterpretation of the second theorem of incompleteness, which states that the consistency of arithmetic cannot be proven within its own formal system. However – Lucas would reply – mathematicians are capable of proving that arithmetic is consistent by making use of more advanced techniques, of languages of a higher order. It is true that the 'going outside the system', which we are capable of doing hardly seems possible for a machine, but what if it could learn to do it? If, from a very complex network of artificial neurons, new visions of consistency emerged? Nothing is as simple as it seems.

What would dear old Euclid make of the bifurcations of the axiomatic method? To disguise him as a 20th century fuzzy logician would be a good end to this novel, which began with the discovery of non-Euclidean geometry, continued with set theory and its paradoxes, and in the following three chapters witnessed the birth of three indisputable heroes: David Hilbert, Kurt Gödel and Alan Turing.

It would be a good ending, but research stays ahead. In the few months it takes for these lines to reach the first readers, mathematicians, physicists and

engineers will have carried on perfecting neural networks, fuzzy logic may well have taken a different direction, and it is even possible that someone may have made progress towards solving the problem 'P versus NP'. Best to leave it like that. *All's Well That Never Ends* is not a bad ending for a book whose protagonists are paradoxes.

Bibliography

DEDEKIND, R., *Essays on the Theory of Numbers,* Whitefish, Kessinger Publishing. 2007.

EUCLIDES, *Elements*, Santa Fe, Green Lion Press, 2002.

HEIJENOORT, J.V., *From Frege to Gödel: A Source Book in Mathematical Logic,* Cambridge (Massachussets), Harvard University Press, 1967.

HOFSTADTER, D.R., *Gödel, Escher, Bach. An Eternal Golden Braid*, London, Penguin. 2000.

JOHNSTONE, P.T., *Notes on Logic and Set Theory*, Cambridge, Cambridge University Press, 1987.

MITCHELL, M., *Complexity*, Oxford, Oxford University Press, 2009.

NAGEL, E. and NEWMAN, J.R., *Gödel's Proof*, New York, New York University Press, 2008.

SANGALLI, A., *The Importance of Being Fuzzy and Other Insights from the Border between Math and Computers*, Princeton, Princeton University Press, 1998.

SMITH, P., *An Introduction to Gödel's Theorems*, Cambridge, Cambridge University Press, 2007

SOKAL, A. and BRICMONT, J., *Intellectual Impostures*, London, Profile Books, 2003.

Index

The Folly of Reason

Mathematical logic and its paradoxes

During the early decades of the 20th century, a pair of eminent logicians, Bertrand Russell and Kurt Gödel, shook the established pillars of the imposing mathematical edifice that had been under careful construction since the times of Euclid. Pioneering the use of logic to understand the foundations of mathematics, some see these figures as a major influence in introducing logic into the absorbing realm of modern computing.